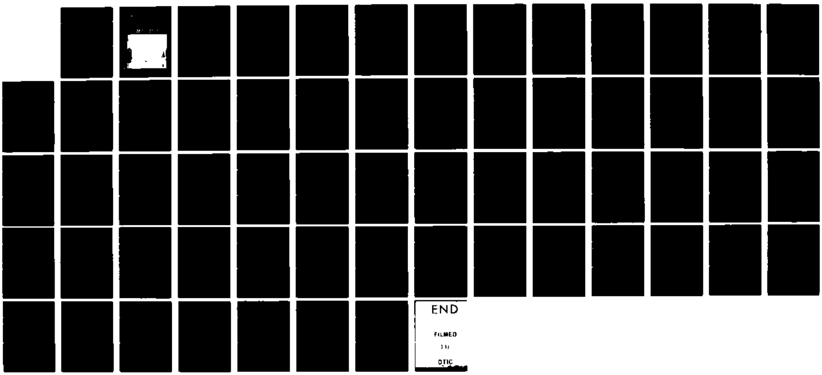
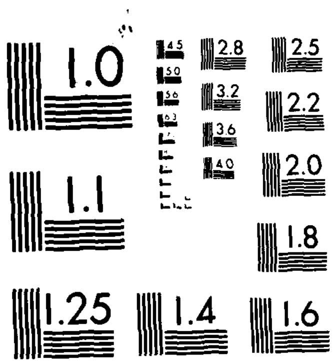
AD-A125 836 ON SEQUENTIAL DETECTION IN DEPENDENT NOISE(U) ILLINOIS UNIV AT URBANA COORDINATED SCIENCE LAB C K CHIANG JUN 80 R-886 N00014-79-C-0424 1/1

UNCLASSIFIED F/G 9/4 NL





MICROCOPY RESOLUTION TEST CHART

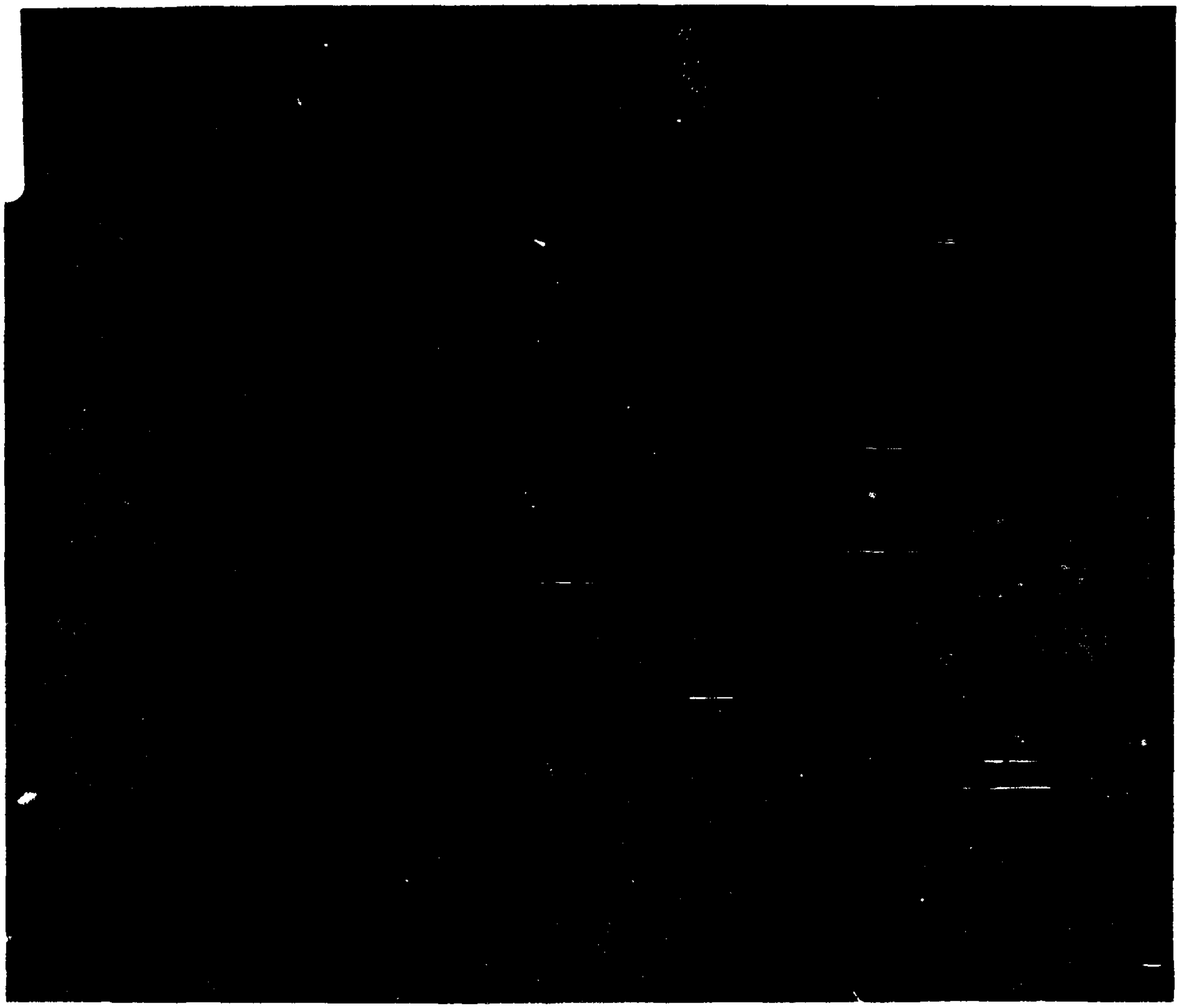
NATIONAL BUREAU OF STANDARDS-1963-A

(15)

REPORT R-886 JUNE, 1980 UILU-ENG 80-2218

COORDINATED SCIENCE LABORATORY

ON SEQUENTIAL DETECTION IN DEPENDENT NOISE



UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ON SEQUENTIAL DETECTION IN DEPENDENT NOISE		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER R-886; UILU-ENG 80-2218
7. AUTHOR(s) Chien Kuo Chiang		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0424
9. PERFORMING ORGANIZATION NAME AND ADDRESS Coordinated Science Laboratory University of Illinois at Urbana-Champaign Urbana, Illinois 61801		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Joint Services Electronics Program		12. REPORT DATE June 1980 13. NUMBER OF PAGES 54
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Sequential Detection
Dependent Noise
Memoryless Detection

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The problem of designing memoryless sequential detection systems for constant signals in m-dependent noise processes is considered. Particular attention is given to the problem of antipodal signaling in an additive noise channel. Applying the criterion of minimum average sample size, the asymptotically optimal detector is shown to be characterized by the solution to a Fredholm integral equation. A paradox arising from the application of standard asymptotic analysis to this detection system is also considered. In addition, a numerical method and resulting solution to the integral equation for the gaussian case is also given.

DD FORM 1473 JAN 73

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ON SEQUENTIAL DETECTION IN DEPENDENT NOISE

by

Chien Kuo Chiang

Supported by the Joint Services Electronics Program (U.S. Army, U.S. Navy U.S. Air Force) under Contract N00014-79-C-0424.

ON SEQUENTIAL DETECTION IN DEPENDENT NOISE

BY

CHIEN KUO CHIANG

B.E., SUNY at Stony Brook, 1978

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1980

Urbana, Illinois

ON SEQUENTIAL DETECTION IN DEPENDENT NOISE

Chien Kuo Chiang
Department of Electrical Engineering
University of Illinois at Urbana-Champaign, 1980

ABSTRACT

The problem of designing memoryless sequential detection systems for constant signals in m-dependent noise processes is considered. Particular attention is given to the problem of antipodal signaling in an additive noise channel. Applying the criterion of minimum average sample size, the asymptotically optimal detector is shown to be characterized by the solution to a Fredholm integral equation. A paradox arising from the application of standard asymptotic analysis to this detection system is also considered. In addition, a numerical method and resulting solution to the integral equation for the Gaussian case is also given.

ACKNOWLEDGEMENTS

The author wishes to express his sincere gratitude and appreciation to Professor H. V. Poor for his constant guidance, encouragement, and advice throughout the course of this work. Thanks are also given to Mrs. Phyllis Young for her excellent typing.

Table of Contents

CHAPTER 1

INTRODUCTION

The field of sequential detection theory was established in the mid 1940's. The early development of this field was based on the work of researchers motivated by wartime needs, such as radar systems.

In more recent years, research on sequential detection theory has been concerned with truncation effects, rank test (P. K. Sen and M. Ghosh 1974, and M. R. Reynolds, Jr. 1975), and mathematical modeling. In the large majority of work in this area the assumption of independent, identically distributed (i.i.d.) observation is used. In this thesis we will concentrate on designing optimum sequential detection procedures for a non-i.i.d. observation model, while retaining the recursive nature of tests designed for i.i.d. models. Because the sample size is not fixed and the samples are not i.i.d., large memory lengths are often required by the detector. Our recursive detector only requires finite memory length so that it is particularly well suited for this purpose.

In Chapter 2, we derive a necessary condition for the optimum recursive detector structure in terms of an integral equation for a nonlinearity characterizing this optimum. A paradox arises in this development and is examined in Chapter 3. Finally a numerical solution of the integral equation is given in Chapter 4.

1.1 The Sequential Probability Ratio Test

Sequential analysis is a method of statistical inference whose characteristic feature is that the number of observations required by the procedure is not determined in advance of the experiment. The decision whether or not to terminate the experiment is made at each stage depending on the results of the observations made up to that point.

The major difference between sequential procedures and conventional fixed sample size (FSS) testing is that the sample size is not fixed in advance. In the conventional fixed sample size (FSS) testing of a simple hypothesis H_0 against a simple alternative H_1, the most powerful test is given by Neyman-Pearson lemma (Van Trees, 1968 page 34).

$$\frac{f_{H_1}(x_1,\ldots,x_n)}{f_{H_0}(x_1,\ldots,x_n)} \underset{H_0}{\overset{H_1}{\gtrless}} T \tag{1.1}$$

where $f_{H_0}(x_1,\ldots,x_n)$ and $f_{H_1}(x_1,\ldots,x_n)$ are the joint densities of the random variables $x_1,\ldots,x_n$ under H_0 and H_1, respectively and where $x_1,\ldots,x_n$ denotes an observation of $X_1,\ldots,X_n$. The sample size n and threshold T are determined by the level of significance desired. If the sample size does not have to be fixed, then a sequential probability ratio test (SPRT) can be defined (A. Wald, 1947) by testing the probability ratio against two thresholds A and B, i.e., we continue sampling as long as

$$A < \frac{f_{H_1}(x_1,\ldots,x_n)}{f_{H_0}(x_1,\ldots,x_n)} < B \tag{1.2}$$

where $0 < A < B < \infty$ and stop when (1.2) is violated. The hypothesis H_0 is accepted if the ratio is less than or equal to A, and the hypothesis H_1 is accepted if the ratio is greater than or equal to B; and the thresholds A and B are chosen to give desired error probabilities. The sample size needed to reach a decision is a random variable defined as

$$N = \min \left\{ n: \frac{f_{H_1}(x_1,\ldots,x_n)}{f_{H_0}(x_1,\ldots,x_n)} \notin (A,B) \right\} \quad . \tag{1.3}$$

Under general conditions $P\{N < \infty\} = 1$. By using Wald's approximation (which assumes the actual stopping position is at one of the thresholds A and B), it can be shown that (A. Wald, 1947, p. 41) we may choose

$$B \cong \frac{\beta}{\alpha}$$

and (1.4)

$$A \cong \frac{1-\beta}{1-\alpha}$$

where β is the desired power and α is the desired size of the test. Thus, via (1.4) one can choose A and B to give desirable probabilities of error. We may also want the error probabilities in terms of boundaries; i.e.,

$$\alpha \cong \frac{1-A}{1-B}$$

(1.5)

$$\beta \cong \frac{B(1-A)}{B-A}$$

It is important to note that these approximations to the error probabilities of a given SPRT(A,B) are independent of the distributions of the data under H_0 and H_1.

1.2 Operating Characteristics and Expected Sample Size Functions

Throughout this section we consider only the independent sampling case. Also, it is assumed that the samples x_i are identically distributed with distribution known except for the values of a finite number of parameters $\theta_1,\ldots,\theta_k$. We use the letter θ without subscript to denote the set of all k parameters $\theta_1,\ldots,\theta_k$. Since the distribution of each x_i is determined by the parameter point θ, the probability of accepting H_0 will be a function of θ. This function is denoted by $L(\theta)$ and is called the operating characteristic (OC) function.

Now, we consider the sequential probability ratio test for testing a simple hypothesis H_0 against a simple alternative H_1. Let $f(x,\theta)$ denote the common probability density function of random variables x_i. Consider the following hypothesis test:

$$\begin{aligned} & H_0\colon\ x_i \sim f(x,\theta_0) \quad \text{all } i \\ \text{vs} & \\ & H_1\colon\ x_i \sim f(x,\theta_1) \quad \text{all } i\ . \end{aligned} \tag{1.5}$$

Consider the expression

$$\left[\frac{f(x,\theta_1)}{f(x,\theta_0)}\right]^{h(\theta)} . \tag{1.6}$$

For each value θ, a value of $H(\theta)$ can be determined such that the expectation of (1.6) is equal to 1, i.e.,

$$\int_{-\infty}^{\infty} \left[\frac{f(x,\theta_1)}{f(x,\theta_0)}\right]^{h(\theta)} f(x,\theta)dx = 1 \tag{1.7}$$

Neglecting the excess of likelihood ratio over the boundaries A and B at the termination of the process, and with Eq. (1.7), an approximated OC function can be shown, to be (Wald, 1947, p. 48),

$$L(\theta) \approx \begin{cases} \dfrac{B^{h(\theta)} - 1}{B^{h(\theta)} - A^{h(\theta)}} \;, & h(\theta) \neq 0 \\[2ex] \dfrac{\text{Log } B}{\text{Log } A + \text{Log } B} \;, & h(\theta) = 0 \end{cases} \tag{1.8}$$

and the expected sample size can be shown to be

$$E_\theta(N) \approx \begin{cases} \dfrac{L(\theta)\text{Log } A + [1-L(\theta)]\text{Log } B}{E_\theta(Z)} \;, & E_\theta(Z) \neq 0 \\[2ex] \dfrac{\text{Log } A \text{ Log } B}{E_\theta(Z^2)} \;, & E_\theta(Z) = 0 \end{cases} \tag{1.9}$$

where $Z = \ell n \left[\dfrac{f(x,\theta_1)}{f(x,\theta_0)}\right]$ and $E_\theta(\cdot)$ denotes expectation with respect to the density $f(x,\theta)$.

1.3 Comparison Between FSS and SPRT

The advantages of SPRT over the FSS are that the thresholds A and B are easier to choose than are T and N. For controlling both probabilities of error under H_0 and H_1, and that the expected sample sizes, $E_{H_0}(N)$ and $E_{H_1}(N)$, under H_0 and H_1, are smaller than the sample size N needed for the same probability of error with FSS test.

There are some drawbacks to the SPRT, one of which is that a more complicated implementation scheme is needed. However, this disadvantage can be compensated for by saving in $E_{H_1}(N)$ and $E_{H_0}(N)$ (A. Wald, 1947, p. 5

gives an example on saving in average sample size of an SPRT compared to the best FSS test of about fifty percent). Even though the SPRT has been shown to terminate with probability one, occasionally a test can go on for a long time before it stops. This excessive test length is usually solved by truncation (see S. Tantartana, 1977). If the test is truncated at a reasonably large sample, then its properties are essentially unaffected by the truncation. One other drawback of the SPRT is that $E_\theta(N)$ can be very large for some values of parameter θ. Under small probabilities of errors, this situation can become worse. Several schemes can reduce $E_\theta(N)$; such as truncating the test, modification of the thresholds A and B (T. W. Anderson, 1960), the minimax scheme (T. L. Lai, 1973), and the generalized SPRT (J. Kieffer and L. Weiss, 1957).

CHAPTER 2

MEMORYLESS SEQUENTIAL DETECTION SYSTEM

The memoryless detector for a constant signal in m-dependent noise under the conventional fixed sample size have been studied previously by H. V. Poor and J. Thomas, (1979). It is shown that, for the case of large samples, an optimum nonlinearity can be obtained by solving the Fredhold integral equation of the second kind, and its driving function is an optimum solution for the independent case. In this chapter, we extend the work of Poor and Thomas to the design of a sequential memoryless detector by using average sample size as the performance measure.

2.1 Problem Definition and Detector Structure

In this section, the problem of designing memoryless sequential detection systems for a constant signal in m-dependent noise processes is considered. A stationary random sequence $\{N\}_{i=1}^{\infty}$ is said to be m-dependent if $\{N_i\}_{i=1}^{\delta}$ and $\{N_i\}_{i=\xi}^{\infty}$ are statistically independent for $\xi \geq \delta \geq 1$ satisfying $\xi - \delta > m$, which is a nonnegative integer. An m-dependent sequence is obtained, for example, by sampling Gaussian processes whose autocorrelation functions have finite support or the output sequences from finite-impulse-response filter with white inputs.

In general, the construction of detectors based on m-dependent noise require a finite memory length. In some cases, it may require unreasonably large lengths of memory if m is large. In this section, we consider the procedures which give the best performance among all memoryless schemes.

To study this problem, we consider detecting a known constant signal in an additive m-dependent noise process. We have a sequence $\{x_i\}_{i=1}^n \equiv x$ of observation of a process $\{X_i\}_{i=1}^n \equiv X$ and we wish to test the antipodal hypothesis,

$$H_0: X_i = N_i - S \qquad i = 1,\ldots,n$$

versus (2.1)

$$H_1: X_i = N_i + S \qquad i = 1,\ldots,n$$

where $\{N_i\}_{i=1}^{\infty}$ is a zero-mean stationary m-dependent noise sequence and S is a known positive constant. We will assume that f, the common univariate density of the noise sequence, is symmetric and strictly positive on the entire real line.

We wish to consider only the memoryless implementation of detection systems for (2.1). A sequential memoryless detector can be represented in the following stopping rule.

$$\varphi(g;x) = \begin{cases} 0 \ ; & \text{if } A < T_g(x) < B \\ 1 \ ; & \text{otherwise} \end{cases} \tag{2.2}$$

where

$$T_g(x) = \sum_{i=1}^{n} g(x_i) \tag{2.3}$$

Here A and B are the boundaries satisfying $0 < A < B < \infty$, and $g(\cdot)$ is a memoryless nonlinearity. The test of Eq. (2.2) and Eq. (2.3) continues sampling as long as $T_g(x)$ stays strictly between A and B, and if $T_g(x) \le A$, hypothesis H_0 is accepted, whereas if $T_g(x) \ge B$, hypothesis H_1 is accepted. The boundaries A and B are chosen to give desired error probability

performance, and $\varphi(g,x)$ is the probability with which we stop. The diagram of the detector structure is shown in Fig. 1, which shows the samples $\{x_i\}_{i=1}^n$ are passed through a nonlinearity g, summed, and then compared to the thresholds.

2.2 Asymptotic Property of T_g

In order to analyze the performance of the detection system described in (2.2), it is necessary to know the asymptotic distribution of the test statistics $T_g(x)$. We may apply a central limit theorem, and under some mild conditions $T_g(x)$ converges to normality under both H_0 and H_1. In particular, we may establish the convergence of the test statistics to a Brownian motion by applying the following theorem of (Billingsley, 1968, p. 174).

Theorem 2.1. (Billingsley, 1968). Let $Y_1, Y_2, \ldots,$ be a φ-mixing sequence of random variables with $\sum_n \varphi_n^{1/2} < \infty$, where $\{\varphi_n\}$ are the mixing coefficients, and assume g is a measurable function satisfying

$$E\{g(Y_1)\} = 0 \quad \text{and} \quad E\{g^2(Y_1)\} < \infty \quad .$$

Define

$$S_n = \sum_{i=1}^{n} g(Y_i) \tag{2.4}$$

and

$$\sigma_0^2 = E\{g^2(Y_1)\} + 2\sum_{j=1}^{\infty} E\{g(Y_1)g(Y_{j+1})\} \tag{2.5}$$

Then, if $\sigma_0^2 > 0$, the sum

$$U_n(t) = \frac{S_{(nt)}}{\sigma_0\sqrt{n}} , \qquad 0 \le t \le 1$$

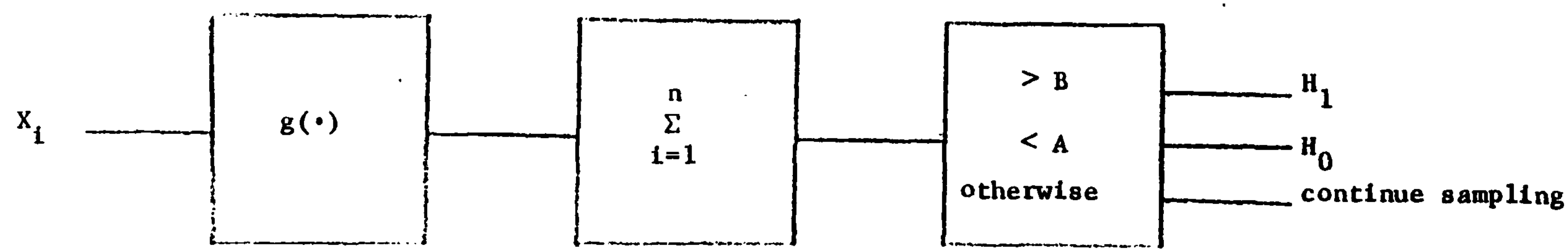


Figure 1. Detector structure.

converges as $n \to \infty$ to a standard Brownian motion with zero mean and variance t in [0,1].

Recall that the definition of a φ-mixing process is given by:

Definition 2.1. Let $\ldots, Y_{-1}, Y_0, Y_1, \ldots$ be a strictly stationary sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. For $a \leq b$, denote $\mathcal{M}_a^b$ as the σ-field generated by the random variables $Y_a, \ldots, Y_b$ (likewise $\mathcal{M}_{-\infty}^a$ is the σ-field generated by $\ldots, Y_{a-1}, Y_a$). Then for a nonnegative function φ of positive integers, we say that the sequence $\{Y_n\}$ is φ-mixing if, for each $k(-\infty < k < \infty)$ and for each $n(n \geq 1)$, $E_1 \in \mathcal{M}_{-\infty}^k$ and $E_2 \in \mathcal{M}_{k+n}^\infty$ together imply

$$|P(E_1 \cap E_2) - P(E_1)P(E_2)| \leq \varphi(n)P(E_1) \tag{2.6}$$

with

$$\lim_{n \to \infty} \varphi(n) = 0.$$

We can see that if $\varphi(n)$ is small, then E_2 is virtually independent of E_1 (weakly dependent). Since an m-dependent sequence ($\varphi(n) = 0$, for $n > m$) is also a φ-mixing sequence, the above theorem automatically holds for an m-dependent sequence. We now restrict the system (2.2) by considering only those detectors based on nonlinearities which satisfy the following mild conditions:

$$E_0\{g(x_1)\} = 0 \tag{2.7a}$$

$$\mathrm{Var}_\theta\{g(X_1)\} < \infty \tag{2.7b}$$

$$|E_\theta\{g(X_1)\}| < \infty \tag{2.7c}$$

and

$$\sigma_\theta^2(g) = \mathrm{Var}_\theta\{g(X_1)\} + 2\sum_{j=1}^{m} \mathrm{Cov}\{g(X_1), g(X_{j+1})\} > 0 \; ; \; \theta = 0,1 \quad \forall\, \theta \in \Theta \tag{2.7d}$$

where the subscript θ denotes expressions computed under H_1 when $\theta = 1$, under H_0 when $\theta = 0$. Under the above conditions, we have from theorem (2.1), that

$$\frac{(T_g(X) - E_\theta\{T_g(X)\})}{\sqrt{n}} \tag{2.8}$$

is asymptotically normally distributed with mean zero and variance $\sigma_\theta^2(g)$.

In other words, $T_g(x)$ is statistically equivalent to $\sum_{i=1}^{n} Y_i$, when Y_i are independent identically distributed with means $\mu_\theta(g)$, and variances $\sigma_\theta^2(g)$. The importance of setting T_g to be a collection of independent observations is that it allows us to study performance (which is the average sample size) under the method of Wald's Fundamental Identity in the following section.

2.3 The Criterion

As we noted in the previous sections, the sequential probability ratio test (SPRT) minimizes average sample size, under hypothesis and alternative. The criterion for our design is the average sample size which already has been approximated by Wald.

In order to proceed with the analysis of the criterion, it is necessary to know the Fundamental Identity of Sequential Analysis (see Ferguson, 1967, p. 373) which stated as follows:

Theorem 2.2. Let $Z_1, Z_2, \ldots$ be independent, identically distributed, finite-valued random variable. Define

$$S_n = \sum_{i=1}^{n} Z_i \tag{2.9}$$

and assume

$$P\{Z_i = 0\} \neq 1, \quad P\{|Z_i| < \infty\} = 1$$

then

$$E\{\exp(tS_N)M(t)^{-N}\} \equiv 1 \tag{2.10}$$

for all t for which M(t) is finite. Here M(t) denotes the moment generating function of the common distribution of Z_i, i.e.,

$$M(t) = E\{\exp(tZ_i)\}, \tag{2.11}$$

and N is defined in Eq. (1.3).

The advantage of using the Fundamental Identity is that we may use Wald's approximation of error probabilities and expected sample size under an arbitrary distribution of the independent identically distributed sequence.

Note that if we can find a nonzero real number t_0 for which $M(t_0) = 1$, then the Fundamental Identity implies

$$E\{\exp(t_0 S_N)\} = 1 \tag{2.12}$$

We may then use (2.12) to approximate the error probabilities by ignoring the excess over the boundaries; that is, (see Ferguson, 1967)

$$\exp(t_0 a)P\{S_N \leq a\} + \exp(t_0 b)P\{S_N \geq b\} \cong 1 \tag{2.13}$$

where $a = \log A$, $b = \log B$ from which we may solve for $P\{S_N \geq b\}$ and $P\{S_N \leq a\}$ by substituting $P\{S_N \geq b\} = 1 - P\{S_N \leq a\}$. We obtain

$$P\{S_N \geq b\} \cong \frac{1 - \exp(t_0 a)}{\exp(t_0 b) - \exp(t_0 a)} \tag{2.14}$$

and

$$P\{S_N \le a\} \cong \frac{\exp(t_0 b) - 1}{\exp(t_0 b) - \exp(t_0 a)} \tag{2.15}$$

Now the approximations for the expected sample size function can be obtained by using Eq. (2.14) and Eq. (2.15) and the following theorem.

Theorem 2.3. (Ferguson, 1967). Assume that $P\{Z_i = 0\} \neq 1$, $P\{|Z_i| < \infty\} = 1$, and that M(t) exist in a neighborhood of the origin. Then

$$\text{(a)} \quad ES_N = \mu EN \tag{2.16}$$

$$\text{(b)} \quad E(S_N - N\mu)^2 = \sigma^2 EN \tag{2.17}$$

where $\mu = EZ_i$ and $\sigma^2 = \text{Var } Z_i$.

Then the approximated sample size function is as follows:

For $\mu_\theta \neq 0$, then

$$E_\theta(N) = \frac{1}{\mu_\theta} E\{S_N\} \cong \frac{1}{\mu_\theta} \{aP\{S_N \le a\} + bP\{S_N \ge b\}\}$$

$$\cong \frac{1}{\mu_\theta} \frac{a\{\exp(t_0 b) - 1\} + b\{1 - \exp(t_0 a)\}}{\{\exp(t_0 b) - \exp(t_0 a)\}} \tag{2.18}$$

For $\mu_\theta = 0$, then

$$E_\theta(N) = \frac{1}{\sigma_\theta^2} E\{S_N^2\} \cong \frac{1}{\sigma_\theta^2} \{a^2 P\{S_N \le a\} + b^2 P\{S_N \ge b\}\} \cong \frac{-ab}{\sigma_\theta^2} \tag{2.19}$$

for all $\theta \in \Theta$, and where σ_θ^2 and μ_θ are the variances and expectations of Z_i under H_0 and H_1.

Applying the above to our case (Eq. (2.3)), where we represent

$$T_g(x) = \sum_{i=1}^{n} Y_i \triangleq T_n$$

with the Y_i Gaussian. The moment generating function of Y_i is given by

$$M(t) = \exp\{t\,\mu(g) + \sigma^2(g)\,\frac{t^2}{2}\} \tag{2.20}$$

where $\mu(\cdot)$ and $\sigma^2(\cdot)$ are the expectation and variance with respect to density function of X. For a non-trivial solution $(t \neq 0)$ of $M(t) = 1$, we get

$$\exp\{t_0\mu(g) + \sigma^2(g)\,\frac{t_0^2}{2}\} = 1 \quad , \quad t_0 \neq 0$$

$$t_0 = -\frac{2\mu(g)}{\sigma^2(g)} \quad . \tag{2.21}$$

Substituting Eq. (2.21) into Eqs. (2.14) and (2.15) we obtain:

$$P\{T_n \geq b\} \cong \frac{1 - \exp\{-\frac{2a\mu(g)}{\sigma^2(g)}\}}{\exp\{-\frac{2b\mu(g)}{\sigma^2(g)}\} - \exp\{\frac{2a\mu(g)}{\sigma^2(g)}\}} \tag{2.22}$$

and

$$P\{T_n \leq a\} \cong \frac{\exp\{-\frac{2b\mu(g)}{\sigma^2(g)}\} - 1}{\exp\{-\frac{2b\mu(g)}{\sigma^2(g)}\} - \exp\{-\frac{2\mu(g)}{\sigma^2(g)}\}} \tag{2.23}$$

Since $P\{T_N \geq b\} = P\{\text{ Choose } H_1\}$ and $P\{T_N \leq a\} = P\{\text{Choose } H_1\}$, we can find approximate error probabilities from (2.22) and (2.23); that is,

$$\beta \cong \frac{1 - \exp\{ - \frac{2a\mu_1(g)}{\sigma_1^2(g)} \}}{\exp\{- \frac{2b\mu_1(g)}{\sigma_1^2(g)} \} - \exp\{- \frac{2a\mu_1(g)}{\sigma_1^2(g)} \}} \tag{2.24}$$

$$\alpha \cong \frac{1 - \exp\{- \frac{2a\mu_0(g)}{\sigma_0^2(g)} \}}{\exp\{- \frac{2b\mu_0(g)}{\sigma_0^2(g)} \} - \exp\{- \frac{2a\mu_0(g)}{\sigma_0^2(g)} \}} \tag{2.25}$$

In view of the i.i.d. optimal detection structure, we may assume that g is an odd-symmetric function about zero. We also assume that the second-order noise density is symmetric. Then we have the following properties:

$$\begin{aligned} \mu_0(g) &= \int_{-\infty}^{\infty} g(x)f(x-s)dx = \int_{-\infty}^{\infty} g(x+s)f(x)dx \\ &= - \int_{-\infty}^{\infty} g(-x-s)f(-x)dx \\ &= - \int_{-\infty}^{\infty} g(x)f(x+s)dx \\ &= - \mu_1(g) \end{aligned} \tag{2.26}$$

and

$$\begin{aligned} \int_{-\infty}^{\infty} g^2(x)f(x-s)dx &= \int_{-\infty}^{\infty} g^2(x+s)f(x)dx \\ &= \int_{-\infty}^{\infty} (-g(-x-s))^2 f(-x)dx \\ &= \int_{-\infty}^{\infty} g^2(x)f(x+s)dx \end{aligned} \tag{2.27}$$

and furthermore, we would like to show $\sigma_1^2(g) = \sigma_0^2(g)$. From Eq. (2.7d) we have

$$\sigma_0^2(g) = \mathrm{Var}_0(g) + 2\sum_{j=1}^{m} \mathrm{Cov}_0(g(X_1),g(X_{j+1})) \tag{2.28}$$

$$\sigma_1^2(g) = \mathrm{Var}_1(g) + 2\sum_{j=1}^{m} \mathrm{Cov}_1(g(X_1),g(X_{j+1})) \tag{2.29}$$

where

$$\mathrm{Var}_\theta(g) = E_\theta(g^2) - E_\theta^2(g) \tag{2.30}$$

substituting Eqs. (2.26) and (2.27) into Eq. (2.30), we find

$$\mathrm{Var}_0(g) = \mathrm{Var}_1(g).$$

Now the problem is equivalent to showing

$$\sum_{j=1}^{m} \mathrm{Cov}_0(g(X_1,g(X_{j+1})) = \sum_{j=1}^{m} \mathrm{Cov}_1(g(X_1),g(X_{j+1})).$$

After some analysis we obtain

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_{j+1})f_{N_1,N_{j+1}}(x_1+s,x_{j+1}+s)dx_1dx_{j+1}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1)g(x_{j+1})f_{N_1,N_{j+1}}(x_1-s,x_{j+1}-s)dx_1dx_{j+1} \tag{2.31}$$

Using the properties of g (that is $\sigma_0^2(g) = \sigma_1^2(g)$ and $-\mu_1(g) = \mu_0(g)$), substituting them in Eq. (2.18), we eliminate the boundaries a and b; and conditioning on H_0 and H_1, we obtain the following expressions:

$$E(N|H_0) \cong \frac{\sigma_0^2(g)}{\mu_0^2(g)} \left[-\tfrac{1}{2}\left(\log(\frac{1-\beta}{1-\alpha})\right)(1-\alpha) - \tfrac{1}{2}\alpha \log(\frac{\beta}{\alpha}) \right] = \frac{\sigma_0^2(g)}{\mu_0^2(g)} K_0 \qquad (2.32)$$

where $K_0 = \tfrac{1}{2}(1-\alpha)\log(\frac{1-\alpha}{1-\beta}) + \tfrac{1}{2}\alpha \log(\frac{\alpha}{\beta})$

and

$$E(N|H_1) \cong \frac{\sigma_1^2(g)}{\mu_1^2(g)} \left[\tfrac{1}{2}(1-\beta)\log(\frac{1-\beta}{1-\alpha}) + \tfrac{1}{2}\beta \log(\frac{\beta}{\alpha}) \right] = \frac{\sigma_1^2(g)}{\mu_1^2(g)} K_1 \qquad (2.33)$$

where $K_1 = \tfrac{1}{2}(1-\beta)\log(\frac{1-\beta}{1-\alpha}) + \tfrac{1}{2}\beta \log(\frac{\beta}{\alpha})$.

We would like to find the optimum nonlinearity g to minimize average sample size under both H_0 and H_1. Assuming g satisfies the constraints defined in previous sections, we see that to minimize $E(N|H_\theta)$ is equivalent to maximizing $\left[\frac{\mu_\theta^2(g)}{\sigma_\theta^2(g)}\right]$, so we may define the performance criterion as follows:

$$S(g) = \left(\frac{\mu_\theta(g)}{\sigma_\theta(g)}\right)^2 \quad , \quad \forall\, \theta \in \Theta \qquad (2.34)$$

Since $\sigma_0^2(g) = \sigma_1^2(g)$ and $-\mu_1(g) = \mu_0(g)$, to minimize $E(N|H_\theta)$ for both $\theta = 0$ and $\theta = 1$ is equivalent to maximizing the criterion $S(g)$ for $\theta = 0$ or 1.

Now a question arises, since the asymtotic approximation of test statistics is used, how accurately can we determine the nonlinearity g? This problem is carefully examined in detail in Chapter 3.

2.4. A Necessary and Sufficient Condition

As we defined in section 2.3, the criterion we used is

$$S(g) = \left\{\frac{(\int g f_\theta)^2}{\sigma_\theta^2(g)}\right\} \quad , \quad \theta = 0,1.$$

where $\sigma_\theta^2(\cdot)$ is the variance as we defined in (2.7d), and f_θ is the univariate observation density under H_0 and H_1.

We can see from Eq. (2.34) that the optimum choice of g_0 will be that g which maximizes the criterion S(g), i.e.,

$$g_0 = \arg\{\max_{g \in \mathcal{G}} S(g)\} \tag{2.35}$$

where $\mathcal{G}$ is the class of g which satisfying the restriction and the assumption we defined in previous sections.

We note that S(g) is invariant to scaling (i.e., $S(g) = S(\alpha g)$ for $\alpha \neq 0$) so that maximizing S(g) is equivalent to maximizing $(\int g f_\theta)^2$ under the constraint that $\sigma_\theta^2(g)$ is equal to a constant. Then Eq. (2.35) is equivalent to

$$g_0 = \arg\{\max_{g \in \mathcal{G}} H(g)\} \tag{2.36}$$

where

$$H(g) = \int g f_\theta + \lambda \sigma_\theta^2(g)$$

and λ is a Lagrange multiplier.

Defining

$$J_g(\epsilon) = H(g + \epsilon \delta g) \tag{2.37}$$

where δg is an arbitrary variation in g, we have a necessary condition for g_0 to solve,

$$J'_{g_0}(0) = 0 , \quad \text{for arbitrary } \delta g . \tag{2.38}$$

We consider Eq. (2.36) under H_1 (which is equivalent to the case under H_0) i.e.,

$$g_0 = \arg\{\max_{g \in \mathcal{G}}\{\int g f(x-s)dx + \lambda\sigma_1^2(g)\}\}$$

consider

$$H(g) = \int g f(x-s)dx + \lambda\sigma_1^2(g) \tag{2.39}$$

We write it in explicit form,

$$H(g) = \int_{-\infty}^{\infty} g(x)f(x-s)dx + \lambda\left[\int_{-\infty}^{\infty} g^2(x)f(x-s)dx + 2\sum_{j=1}^{m}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1)g(x_{j+1})\right.$$

$$\left. f_{N_1,N_{j+1}}(x_1-s,x_{j+1}-s)dx_1dx_{j+1} - (2m+1)\left(\int_{-\infty}^{\infty} g(x)f(x-s)dx\right)^2\right] \tag{2.40}$$

We now use the symmetry properties of g and f, rewrite it in the following form

$$H(g) = \int_0^{\infty} g(x)f(x-s)dx - \int_0^{\infty} g(x)f(x+s)dx + \lambda\left[\int_0^{\infty} g^2(x)f(x-s)dx + \int_0^{\infty} g^2(x)f(x+s)dx\right.$$

$$+ 2\sum_{j=1}^{m}\left\{\int_0^{\infty}\int_0^{\infty} g(x_1)g(x_{j+1})f_{N_1,N_{j+1}}(x_1-s,x_{j+1}-s)dx_1dx_{j+1} - \int_0^{\infty}\int_0^{\infty} g(x_1)\right.$$

$$g(x_{j+1})f_{N_1,N_{j+1}}(x_1-s,x_{j+1}+s)dx_1dx_{j+1} - \int_0^{\infty}\int_0^{\infty} g(x_1)g(x_{j+1})f_{N_1,N_{j+1}}$$

$$(x_1+s,x_{j+1}-s)dx_1dx_{j+1} + \int_0^{\infty}\int_0^{\infty} g(x_1)g(x_{j+1})f_{N_1,N_{j+1}}(x_1+s,x_{j+1}+s)$$

$$\left.\left. dx_1dx_{j+1}\right\} - (2m+1)\left(\int_0^{\infty} g(x)f(x-s)dx - \int_0^{\infty} g(x)f(x+s)dx\right)^2\right] \tag{2.41}$$

then substitute $g = g + \epsilon\delta g$ in Eq. (2.41) and take derivative with respect to ϵ. We obtain,

$$\frac{\partial H(g+\epsilon g)}{\partial\epsilon} = \int_0^\infty \delta g f(x-s)dx - \int_0^\infty \delta g f(x+s)dx + \lambda\Big[2\int_0^\infty (g+\epsilon\delta g)\delta g f(x-s)dx + 2$$

$$\int_0^\infty (g+\epsilon\delta g)\delta g f(x+s)dx + 2\sum_{j=1}^{m}\Big\{2\int_0^\infty\int_0^\infty (g+\epsilon\delta g)\delta g f_{N_1,N_{j+1}}(x_1-s,x_{j+1}-s)$$

$$dx_1dx_{j+1} - 2\int_0^\infty\int_0^\infty (g+\epsilon\delta g)\delta g f_{N_1,N_{j+1}}(x_1-s,x_{j+1}+s)dx_1dx_{j+1} - 2\int_0^\infty\int_0^\infty$$

$$(g+\epsilon\delta g)\delta g f_{N_1,N_{j+1}}(x_1+s,x_{j+1}-s)dx_1dx_{j+1} + 2\int_0^\infty\int_0^\infty (g+\epsilon\delta g)\delta g f_{N_1,N_{j+1}}$$

$$(x_1+s,x_{j+1}+s)dx_1dx_{j+1}\Big\} - 2(2m+1)\left(\int_0^\infty \delta g f(x-s)dx - \int_0^\infty \delta g f(x+s)dx\right)$$

$$\left(\int_0^\infty (g+\epsilon\delta g)f(x-s)dx - \int_0^\infty (g+\epsilon\delta g)f(x+s)dx\right)\Big] \tag{2.42}$$

Then we set $\epsilon = 0$, and after some manipulations we have

$$J'_g(0) = \int_0^\infty \Big\{f(x-s)-f(x+s)+2\lambda[g(x)f(x-s)+g(x)f(x+s)+2\sum_{j=1}^{m}\Big\{\int_0^\infty g(x_{j+1})f_{N_1,N_{j+1}}$$

$$(x_1-s,x_{j+1}-s)dx_{j+1} - \int_0^\infty g(x_{j+1})f_{N_1,N_{j+1}}(x_1-s,x_{j+1}+s)dx_{j+1} - \int_0^\infty g(x_{j+1})$$

$$f_{N_1,N_{j+1}}(x_1+s,x_{j+1}-s)dx_{j+1} + \int_0^\infty g(x_{j+1})f_{N_1,N_{j+1}}(x_1+s,x_{j+1}+s)dx_{j+1}\Big\}$$

$$-(2m+1)(f(x-s)\int_0^\infty f(u-s)g(u)du - f(x-s)\int_0^\infty g(u)f(u+s)du - f(x+s)\int_0^\infty g(u)f(u-s)$$

$$du + f(x+s)\int_0^\infty g(u)f(u+s)du)]\Big\}\delta g(x)dx \tag{2.43}$$

where $f_{N_1,N_{j+1}}$ is the joint probability density function of N_1 and N_{j+1}. Since δg is arbitrary, $J_g'(0)$ will be zero, if and only if, the quantity in the brackets $\{\ldots\}$ is identically equal to zero. After arrangement we have,

$$\hat{g}(x)+2\lambda\int_0^\infty K(x,y)g(y)dy = -2\lambda g(x) \qquad \forall\ x \in (-\infty,\infty) \tag{2.44}$$

where

$$\hat{g}(x) = \frac{L(x)-1}{L(x)+1} \tag{2.45}$$

$$K(x,y) = 2\left\{\left\{\sum_{j=1}^{m}(f_{N_1,N_{j+1}}(x-s,y-s)-f_{N_1,N_{j+1}}(x-s,y+s)f_{N_1,N_{j+1}}(x+s,y-s) + f_{N_1,N_{j+1}}(x+s,y+s))/(f(x-s)+f(x+s))-Q(x,y)\right\}-Q(x,y)\right\} \tag{2.46}$$

where

$$Q(x,y) = \hat{g}(x)(f(y-s)-f(y+s)) \tag{2.47}$$

and

$$L(x) = \frac{f(x-s)}{f(x+s)} \tag{2.48}$$

Here $\hat{g}(x)$ is a driving function and $K(x,y)$ is the kernel.

A sufficient condition for g_0 to maximize $S(g)$ is that $Jg_0(\epsilon) \leq Jg_0(0)$ for arbitrary δg and ϵ. We have (details see Appendix)

$$J_g(\epsilon) = J_g(0) + \epsilon J_g'(0) + \lambda(\epsilon^2\sigma_1^2(\delta g)) \tag{2.49}$$

If g_0 satisfies (2.44), we must have

$$J_{g_0}(\epsilon) = J_{g_0}(0) + \lambda(\epsilon^2\sigma_1^2(\delta g)) \leq J_{g_0}(0)$$

for all ϵ and δg, we conclude that λ is negative. Then with negative λ is both necessary and sufficient for g_0 to maximize $S(g)$.

Since λ is arbitrary (but negative), we may choose $\lambda = -\frac{1}{2}$ by analogy with the case in which the kernel is identically zero (equivalent to the independent-sampling case), we then have (2.44) in the form

$$\hat{g}(x) - \int_0^\infty K(x,y)g(y)dy = g(x) \tag{2.50}$$

which is a Fredholm equation of the second kind. In Chapter 4 we give the solution for this integral equation.

CHAPTER 3

COMPARISON BETWEEN TANH($\frac{\text{LogL}}{2}$) and Log L

In this section, the problem of a paradox which arises from the development of standard asymptotic analysis to sequential signal detection systems is considered. The asymptotic analysis of the test statistics leads to an optimum nonlinearity which seems to contradict the well-known Wald-Wolfowitz theorem. Moreover, this situation can be extended to general detection systems. In other words, an analogous paradox exists in the likelihood-ratio process which seems to contradict the Neyman-Pearson lemma and Baysian theory.

3.1 The Optimum Nonlinearity TANH((LogL)/2)

Consider Eq. (4.7), if we set m = 0, in which kernel is identically zero, then it reduces to the independent-sampling case. The solution for the optimum nonlinearity g_0 of this integral equation is proportional to TANH((LogL)/2) (which can also be written as (L-1)/(L+1)). When L(x) is defined in Eq. (4.11) this seems to contradict the well-known Wald and Wolfowitz theorem (see A. Wald 1947) which proves that Log(L) is the optimum nonlinearity function that minimizes the average sample size under both H_0 and H_1. Now, it is necessary to show that our nonlinearity TANH((LogL)/2) is <u>the</u> optimum solution for the independent-sampling case, under our criterion S(g).

Consider our new detection hypothesis:

$$H_0 : X_i = N_i - s \qquad i = 1,\ldots,h$$

vs.

$$H_1 : X_i = N_i + s \qquad i = 1,\ldots,h \tag{3.1}$$

where N_i is a zero mean i.i.d. noise random sequence, and f_N is the common univariate, even symmetric noise density, s is a known positive number. Again assume g is an odd symmetric function about zero, then we have the properties:

$$\mu_0(g) = \int_{-\infty}^{\infty} g(x)f_N(x+s)dx = \int_{-\infty}^{\infty} g(x)f_N(x-s)dx = \mu_1(g) \tag{3.2}$$

and

$$\mu_0(g^2) = \int_{-\infty}^{\infty} g^2(x)f_N(x+s)dx = \int_{-\infty}^{\infty} g^2(x)f_N(x-s)dx = \mu_1(g^2)$$

or

$$\sigma_1^2(g) = \sigma_0^2(g) \tag{3.3}$$

Similarly the sample size that is

$$E_\theta(N) = \frac{\sigma_\theta^2(g)}{\mu_\theta^2(g)} K_\theta , \qquad \forall\, \theta \in \Theta \tag{3.4}$$

and again our criterion is

$$S(g) = \left[\frac{\mu_\theta(g)}{\sigma_\theta(g)}\right]^2 , \qquad \forall\, \theta \in \Theta \tag{3.5a}$$

Again, to maximize S(g) under H_1 is equivalent to maximize S(g) under H_0 so that we can write

$$S(g) = \left[\frac{\mu_1(g)}{\sigma_1(g)}\right]^2 \tag{3.5b}$$

Now, we would like to show that the g_0 = TANH((LogL)/2) is the function maximizing S(g) for the independent sampling case.

Consider the following theorem:

<u>Theorem 3.1</u>. Suppose $\mathcal{J}$ is the class of all odd-symmetric nonlinearities satisfying

$$\mu_1(g^2) < \infty .$$

Then

$$\arg\left\{ \max_{g \in \mathcal{J}} \left\{ \frac{\mu_1^2(g)}{\sigma_1^2(g)} \right\} \right\} = \frac{L-1}{L+1} = \mathrm{TANH}\left(\frac{\mathrm{Log} L}{2}\right) \tag{3.6}$$

where L is the single sample likelihood ratio defined by

$$L(x) = \frac{f_N(x-s)}{f_N(x+s)}$$

<u>Proof</u>. Suppose $\mu_1(g)$ is not identically zero, then $\left[\frac{\mu_1(g)}{\sigma_1(g)}\right]^2$ will have a maximum over $\mathcal{J}$ and we can write

$$\frac{\mu_1^2(g)}{\sigma_1^2(g)} = \frac{\mu_1^2(g)}{\mu_1(g^2) - \mu_1^2(g)}$$

$$= \frac{1}{\left\{\frac{\mu_1(g^2)}{\mu_1^2(g)}\right\} - 1}$$

so that maximizing $\frac{\mu_1^2(g)}{\sigma_1^2(g)}$ is equivalent to maximizing

$$W(g) \triangleq \frac{\mu_1^2(g)}{\mu_1(g^2)}$$

To write the above equation in explicit form, we have

$$W(g) = \frac{\left\{\int_{-\infty}^{\infty} g(x) f_N(x-s)\,dx\right\}^2}{\int_{-\infty}^{\infty} g^2(x) f_N(x-s)\,dx} \tag{3.7}$$

Since g(x) is, by hypothesis, an odd function of x, and $g^2(x)$ is an even function of x, we have

$$\mu_1(g) = \int_{-\infty}^{\infty} g(x) f_N(x-s)\,dx = \int_0^{\infty} g(x) f_N(x-s)\,dx - \int_0^{\infty} g(x) f_N(x+s)\,dx$$

$$= \int_0^{\infty} \left\{ g(x) \left(\frac{L(x)-1}{L(x)+1} \right) \right\} \cdot [f_N(x-s)+f_N(x+s)]\,dx \tag{3.8}$$

and

$$\mu_1(g^2) = \int_{-\infty}^{\infty} g^2(x) f_N(x-s)\,dx = \int_0^{\infty} g^2(x) f_N(x-s)\,dx + \int_0^{\infty} g^2(x) f_N(x+s)\,dx$$

$$= \int_0^{\infty} g^2(x) [f_N(x-s) + f_N(x+s)]\,dx \quad . \tag{3.9}$$

Substituting (3.9) and (3.8) in (3.7), we then have

$$W(g) = \frac{\left\{ \int_0^{\infty} \left\{ g(x) \cdot \frac{L(x)-1}{L(x)+1} \right\} [f_N(x-s) + f_N(x+s)]\,dx \right\}^2}{\int_0^{\infty} g^2(x) [f_N(x-s) + f_N(x+s)]\,dx} \tag{3.10}$$

Applying Schwarz's inequality, i.e.,

$$\left(\int q(x)h(x)dx\right)^2 \le \int q^2(x)dx \int h^2(x)dx$$

to the numerator of (3.10) where

$$q(x) \triangleq g(x)[f_N(x-s) + f_N(x+s)]^{\frac{1}{2}}$$

$$h(x) \triangleq \frac{L(x)-1}{L(x)+1}\,[f_N(x-s) + f_N(x+s)]^{\frac{1}{2}}$$

then

$$W(g) \le \int_0^\infty \left[\frac{L(x)-1}{L(x)+1}\right]^2 [f_N(x-s) + f_N(x+s)]dx \tag{3.11}$$

with equality, if and only if $g_0(x) = \gamma\left(\frac{L(x)-1}{L(x)+1}\right)$. It can be easily shown that $\frac{L(x)-1}{L(x)+1}$ is an odd symmetric function, which agrees with our previous assumption. For some real $\alpha \neq 0$, we have the optimum nonlinearity, i.e.,

$$\frac{L(x)-1}{L(x)+1} = \mathrm{TANH}\left(\frac{\mathrm{LogL}}{2}\right)$$

Q.E.D.

From the above proof, we can say that $\mathrm{TANH}\left(\frac{\mathrm{LogL}}{2}\right)$ indeed is the optimum nonlinearity under our criterion. Now a paradox arises since $\mathrm{TANH}\left(\frac{\mathrm{LogL}}{2}\right)$ and LogL are both derived to be optimum under the same hypothesis. To resolve this paradox, we must trace back to Chapter 2. Because if we assume asymptotic normality of the detection statistics and use the approximated average sample size as our criterion, then it is understandable why they are different, since LogL is derived under the actual expression of sample size and without using asymptotic approximation properties of the test statistics. However, in general for the case of large sample size, asymptotic approximations for the test statistics are often

accurate and the approximated criterion are widely used and acceptable. Since they are different, we also would like to compare their performance which is the average sample size, under a given test. The details and numerical results are presented in the next section.

3.2 Numerical Results and Examples

In this section, we compute the sample size for the two nonlinearities under different noise densities. We also use signal (s), variance (σ^2) and error probabilities (α and β) as our parameters.

Several noise densities are examined;

(a) Generalized Gaussian noise:

This probability density function is defined by

$$f_c(x) = [c\eta(\sigma,c)/(2\Gamma(1/c))]\exp\{-[\eta(\sigma,c)|x|]^c\} \tag{3.12}$$

where

$$\eta(\sigma,c) \triangleq \sigma^{-1}[\Gamma(3/c)/\Gamma(1/c)]^{\frac{1}{2}}.$$

Here c is a positive parameter controlling the rate of decay, $\Gamma(\cdot)$ is the gamma function, and σ^2 is the noise variance. Note that for c = 2 this density reduces to the Gaussian density, whereas for c = 1 it becomes the Laplace density. The density of (3.12) is illustrated in Fig. 2 for values of $\sigma^2 = 1$, and c = 1.5, 2, and 3. Now, we substitute Eq. (3.12) in Eq. (3.6) to obtain

$$L(x) = \frac{f_c(x-s)}{f_c(x+s)} = \exp\{-\frac{\gamma^c(c)}{2}(|x-s|^c - |x+s|^c)\} \tag{3.13}$$

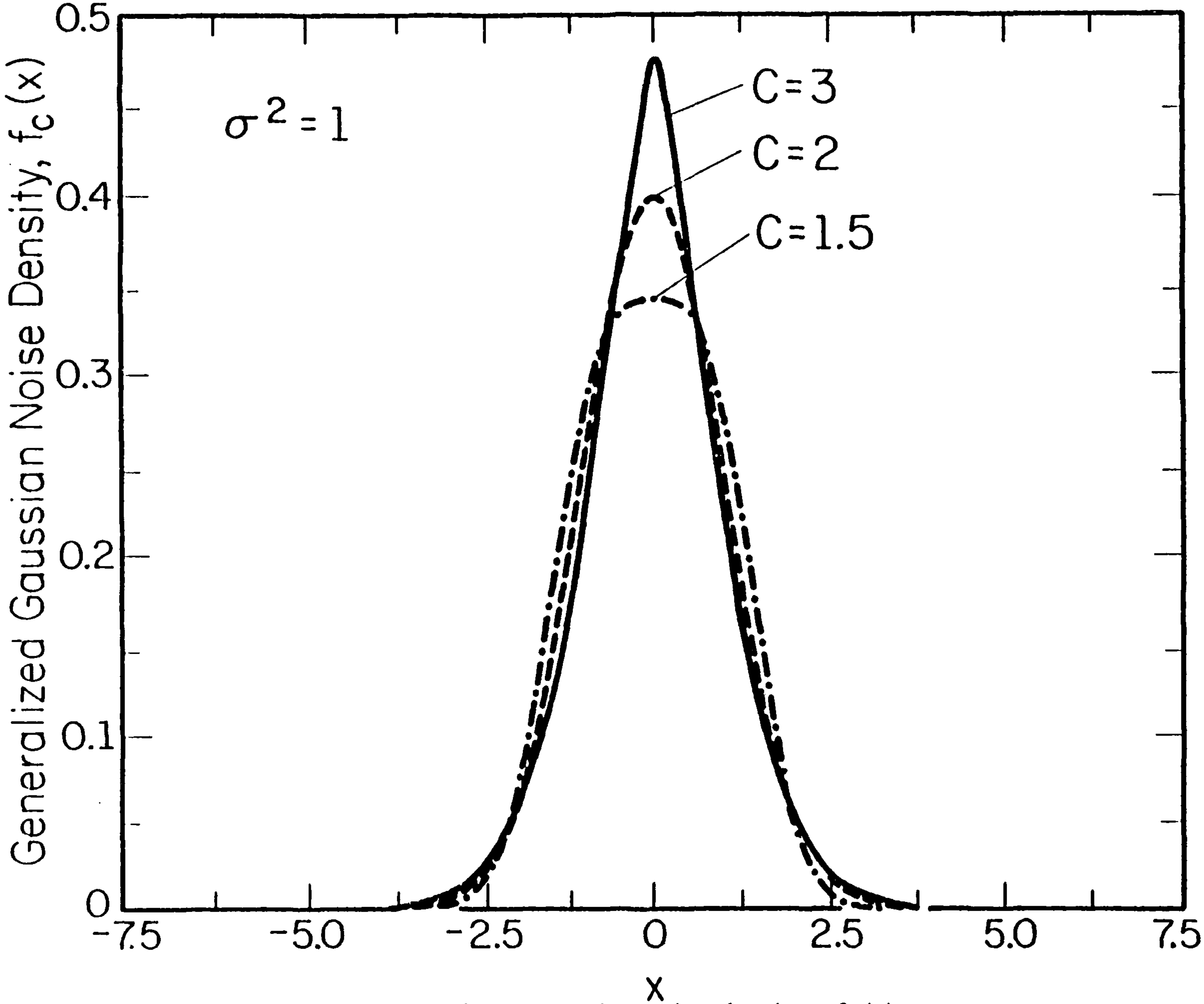


Figure 2. Generalized Gaussian noise density, $f_c(x)$.

and

$$\mathrm{TANH}\left(\frac{\mathrm{LogL}}{2}\right) = \mathrm{TANH}\left\{-\frac{\gamma^c(c)}{2}\left(|x-s|^c - |x+s|^c\right)\right\} \tag{3.14}$$

where

$$\gamma^c(c) = \left[\frac{\Gamma(3/c)}{\Gamma(1/c)}\right]^{\frac{1}{2}} .$$

Substituting Eq. (3.13) and Eq. (3.14) in Eq. (3.4), then we approximate $E_1(N)$. Also for $c=2$, the nonlinearities are plotted in Fig. 9. The sample size ratio between TANH $\left(\frac{\mathrm{LogL}}{2}\right)$ and LogL under H_1 is defined as follows

$$\mathrm{Ratio} \triangleq R \triangleq \frac{\hat{E}_1[N|\mathrm{TANH}(\mathrm{LogL}/2)]}{\hat{E}_1[N|\mathrm{LogL}]} \tag{3.15}$$

where $\hat{E}_1$ denotes approximate sample size computed by (3.4). Note that R is a function of s, α, β, c, and σ^2. For fixed α, β, and $\sigma^2=1$, the plot of R versus c under different values of signal strength is shown in Fig. 3. Also for $c=2$ and $\sigma^2=1$, the plot of TANH $\frac{\mathrm{LogL}}{2}$ versus logL is shown in Fig. 4.

The results show, as it might be expected, that, according to the approximation, the nonlinearity TANH(LogL/2) requires fewer samples than LogL requires, and this difference becomes significant as c and signal strength increases. We may infer that under small signals, the nonlinearity TANH(logL/2) performs approximately as well as logL does. From Fig. 4, we can see that for the regions between -5 and 5, LogL and TANH(LogL/2) are essentially the same.

(b) <u>Generalized Cauchy Noise</u>

$$f_c(x) = [c\eta(\sigma,c)/2\Gamma(1/c)][\nu^{-1/c}\Gamma(\nu+\tfrac{1}{c})/\Gamma(\nu)]\cdot$$
$$\cdot\{1+[\eta(\sigma,c)|x|]^c/\nu\}^{-(\nu+\frac{1}{c})} \tag{3.16}$$

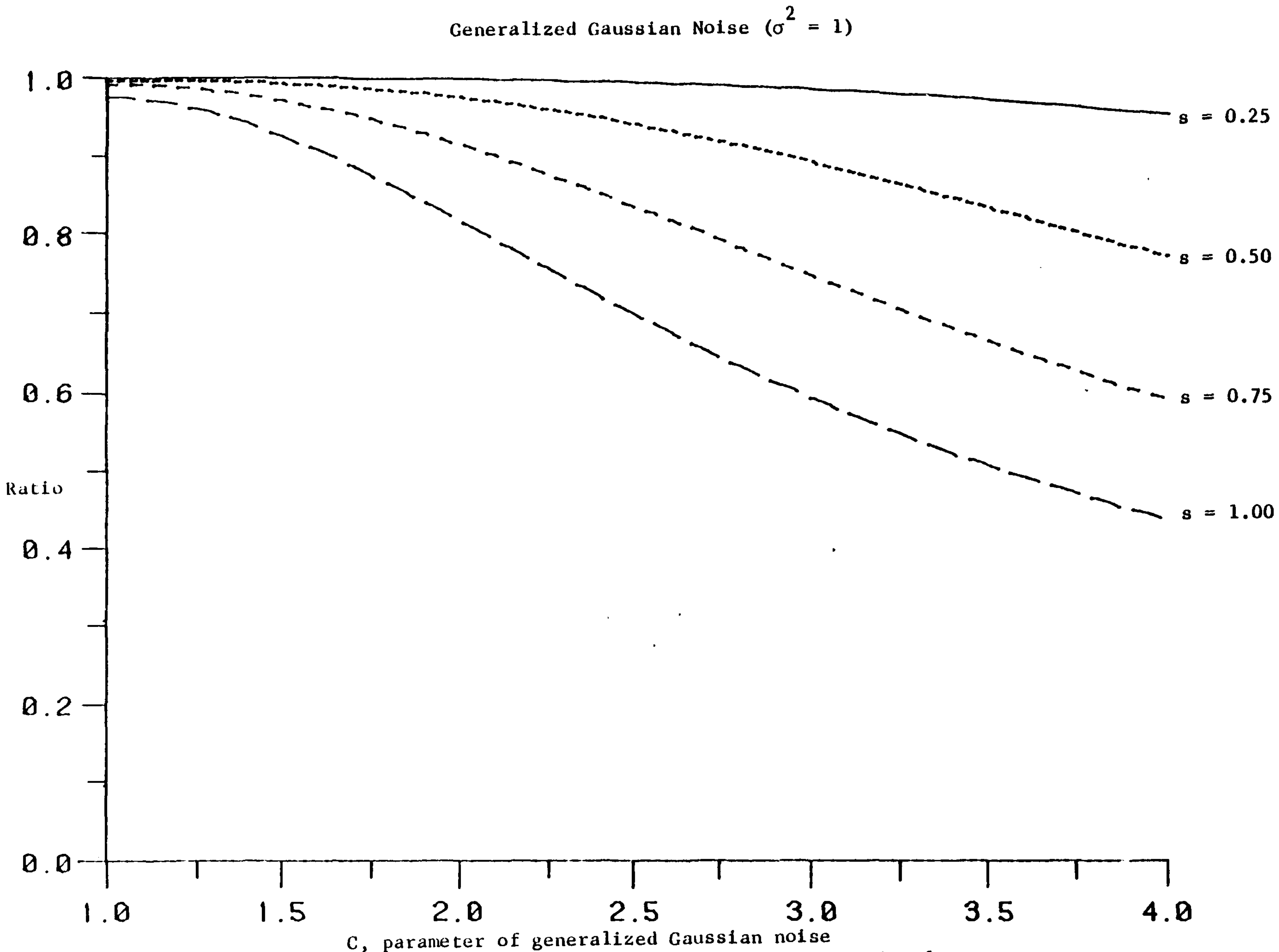


Figure 3. Sample ratio under different signals.

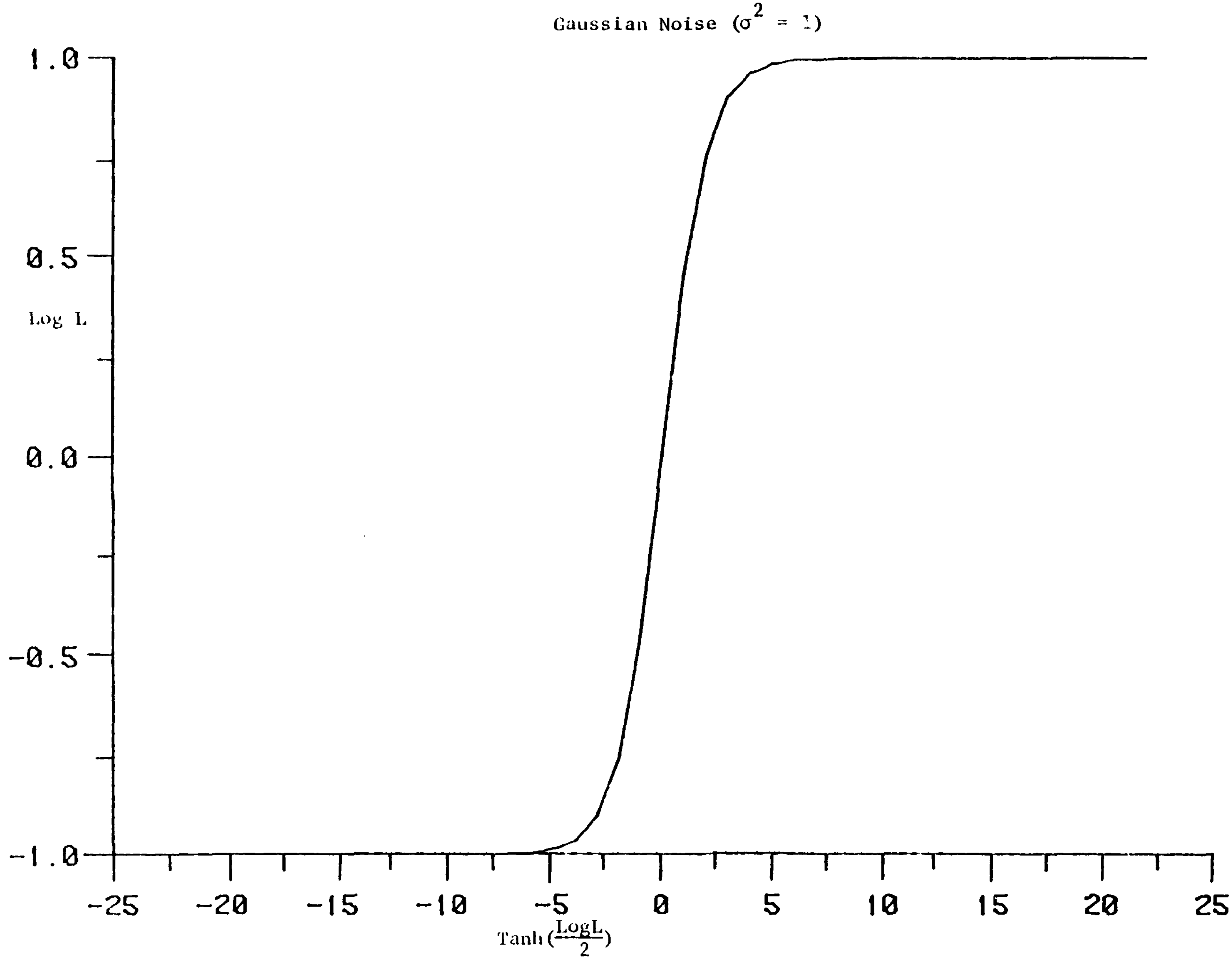


Figure 4. Nonlinearities (relation ratio) for Gaussian noise.

where c and ν are positive parameters. Note that for $c = 2$ and $\nu = \frac{1}{2}$ we have the Cauchy density. We also know that as the parameter ν in the distribution of (3.16) approaches infinity, the resulting distribution approaches the generalized Gaussian distribution of (3.12). However unlike the generalized Gaussian density which always has a variance, the variance of the generalized Cauchy density will be finite only if $c\nu > 2$ and for all our computations $\nu = \frac{1}{2}$ is assumed. For $c = 2$, the nonlinearities are plotted in Fig. 10 and Fig. 6. The plot of the approximate sample ratio versus c is shown in Fig. 5, again the ratio decreases as c and signal increases, and surprisingly, for the case $c = 4$ and signal $= 1$, the ratio still has 0.994.

(c) Hyperbolic Secant Noise

$$f_s(x) = [(\exp\{\pi x/2\sigma\} + \exp\{-\pi x/2\sigma\})]^{-1}$$

$$= \mathrm{Sech}(\pi x/2\sigma)/(2\sigma) \quad . \tag{3.17}$$

Similar graphs are plotted in Fig. 11, Fig. 7 and Fig. 8. From that, we may conclude that the sample ratio always decreases as signal increases. In Table I and Table II, are given the sample sizes under different signal and error probabilities. We can see for the same condition, our approximation implies that Cauchy noise requires more samples than Gaussian noise or hyperbolic secant noises require, mainly because the variance is not limited under Cauchy noise.

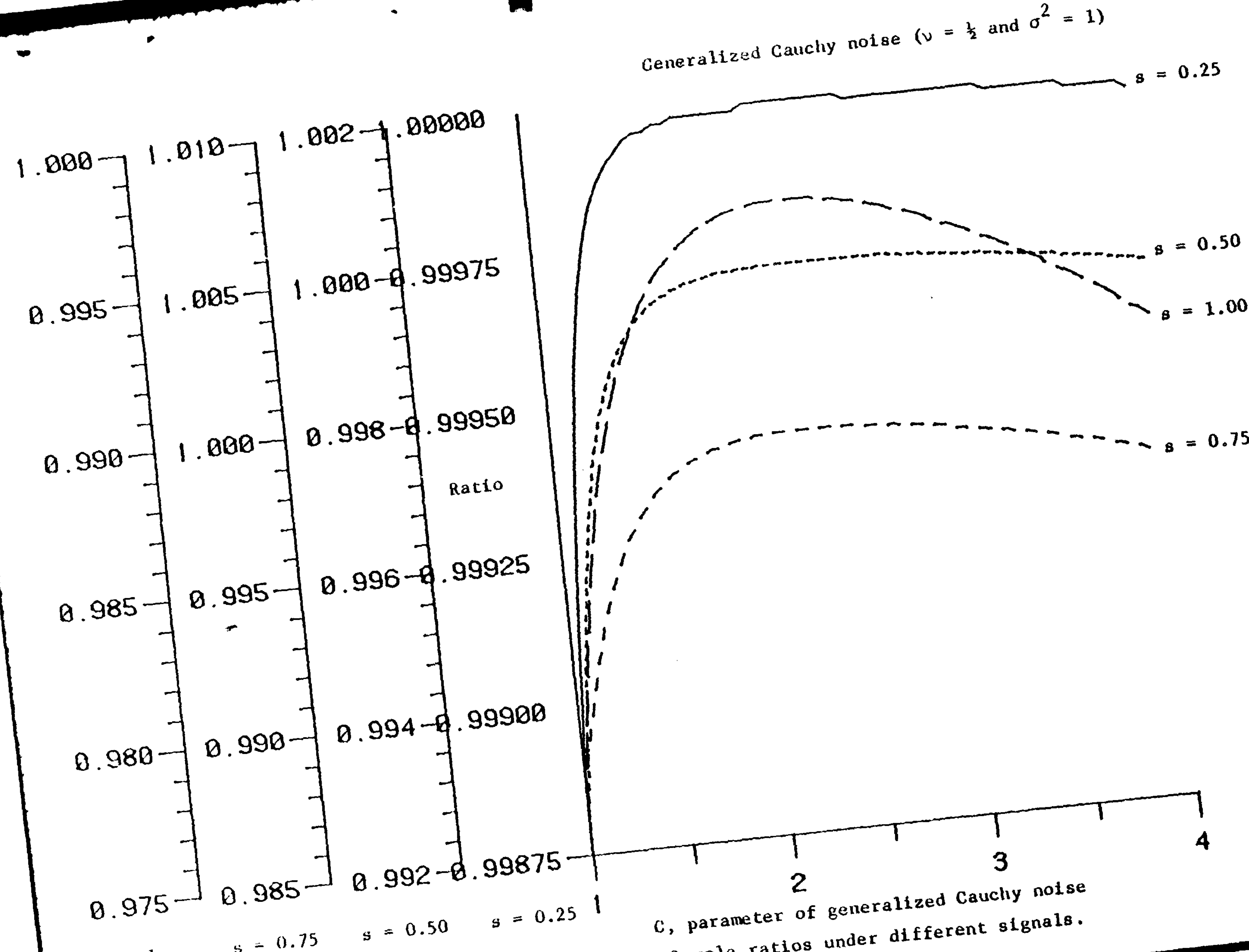


Figure 5. Sample ratios under different signals.

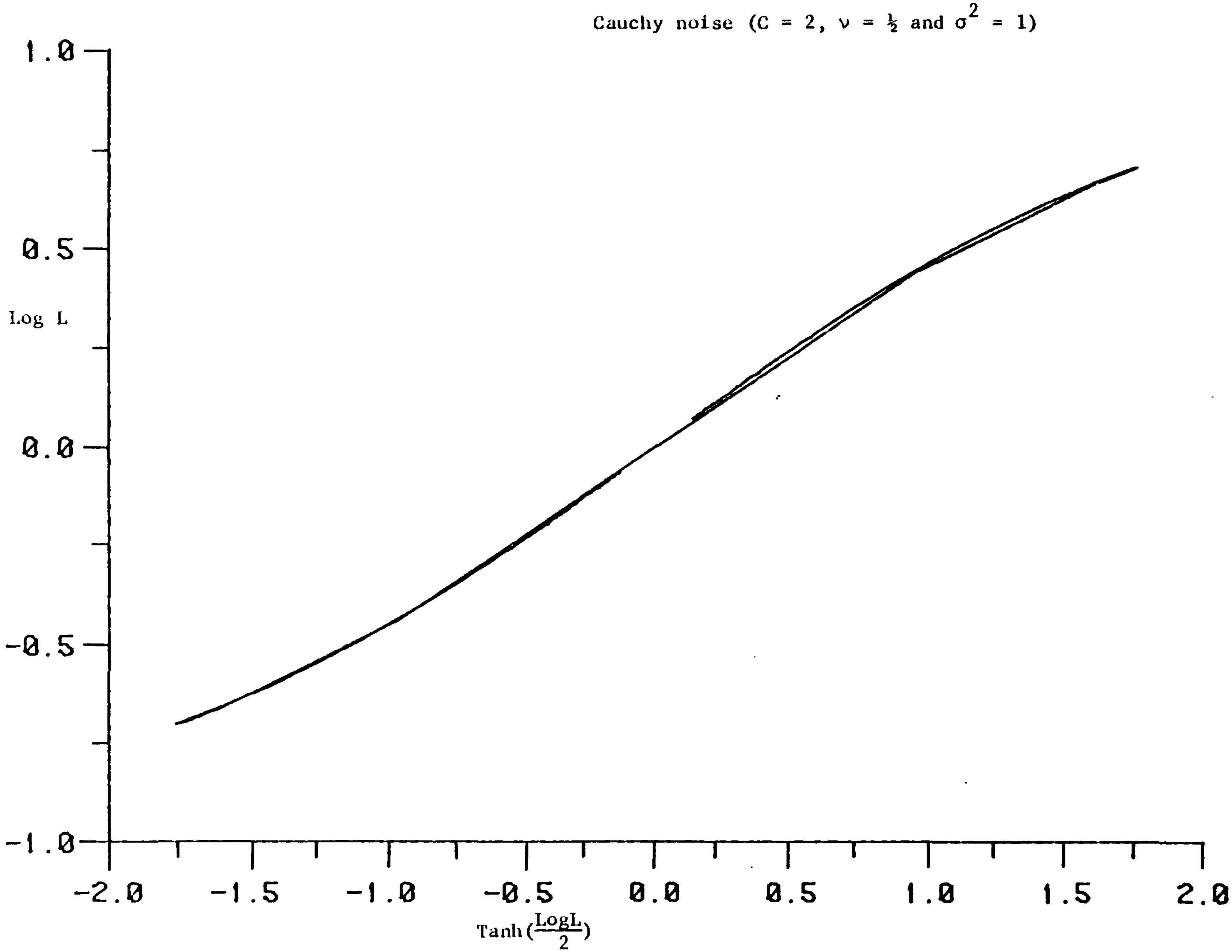


Figure 6. Nonlinearities (relative ratio) for Cauchy noise.

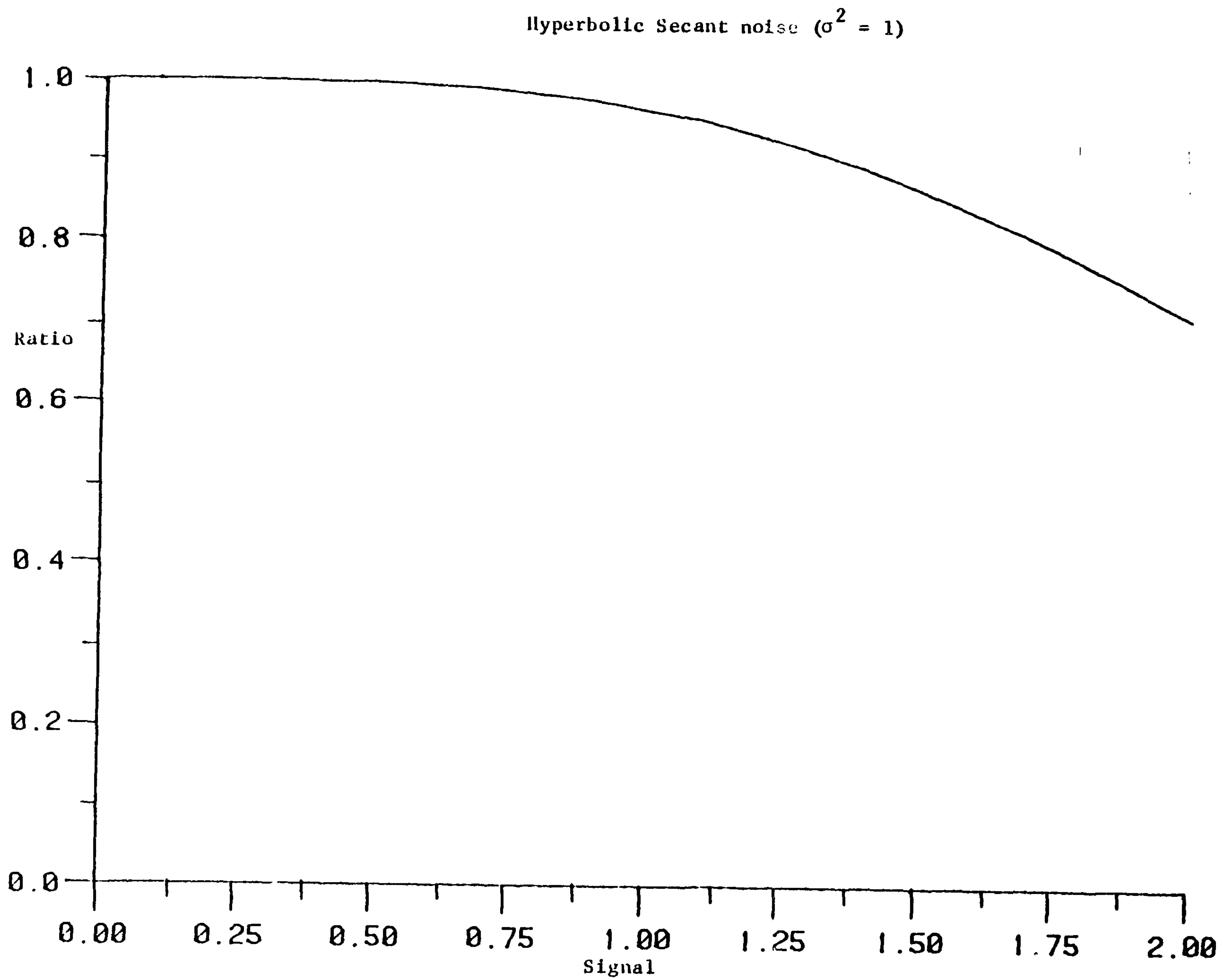


Figure 7. Sample ratio versus signals under Sech noise.

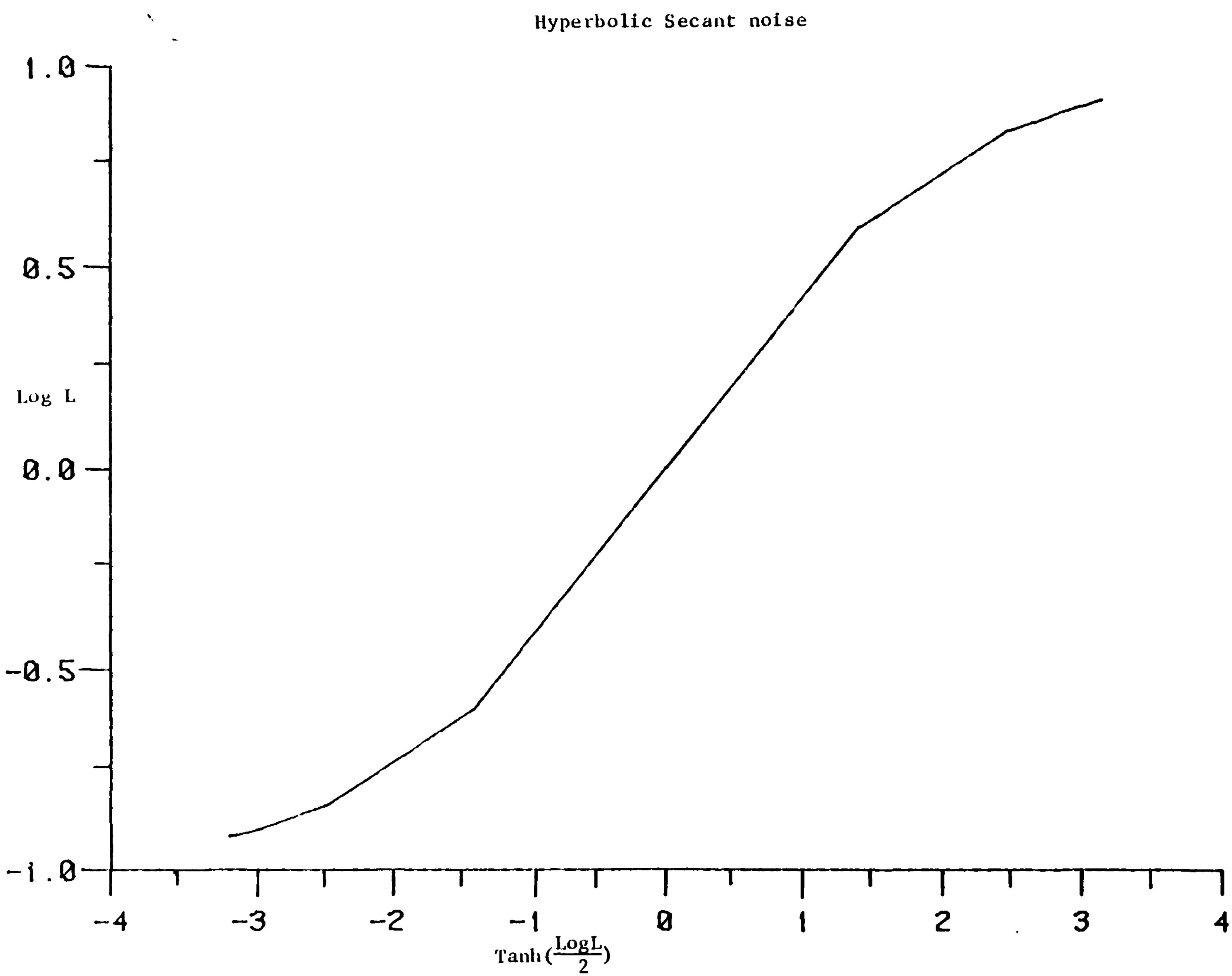


Figure 8. Nonlinearities (relative ratio) for Sech noise.

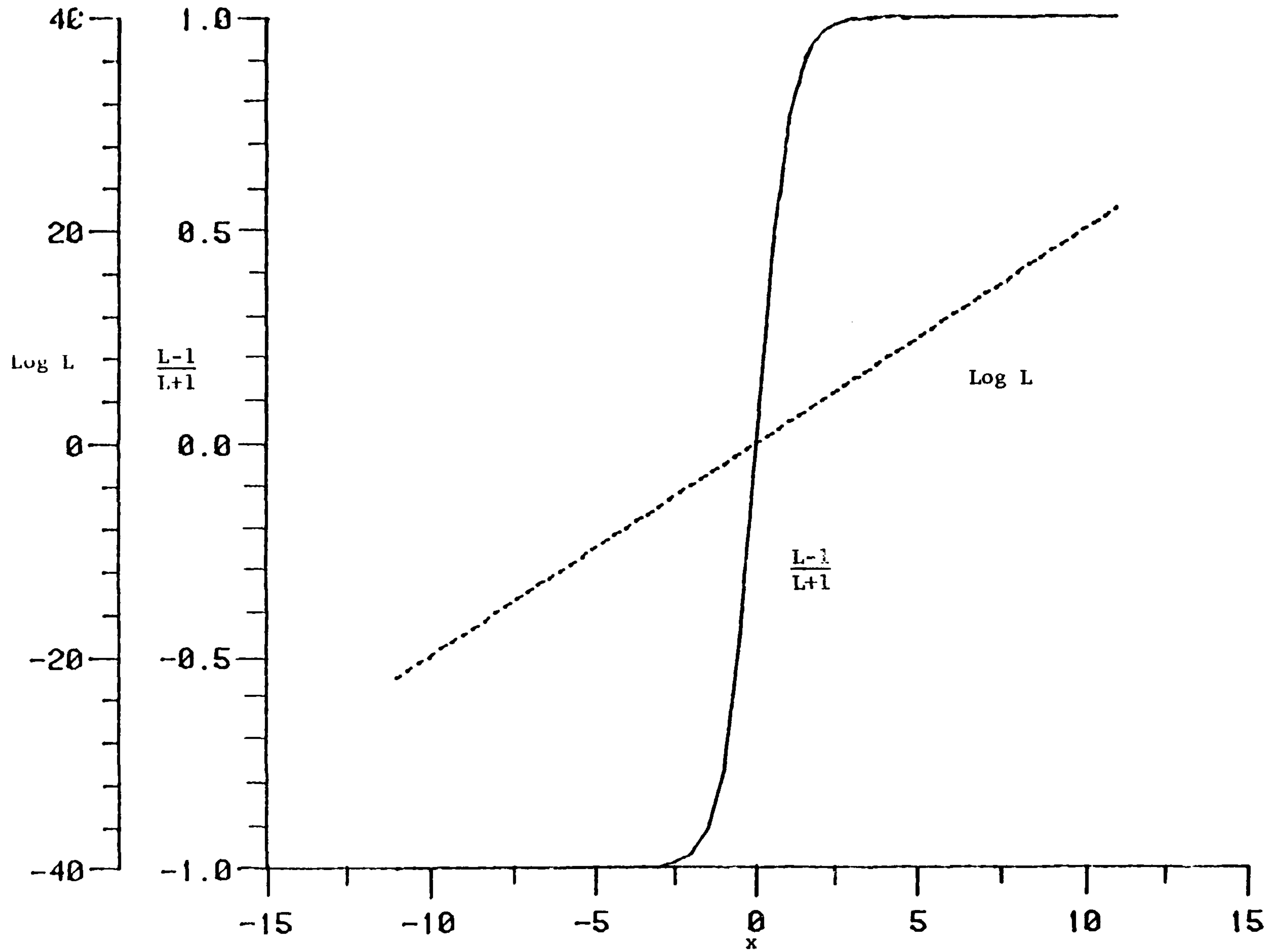


Figure 9. Nonlinearities for the Gaussian noise.

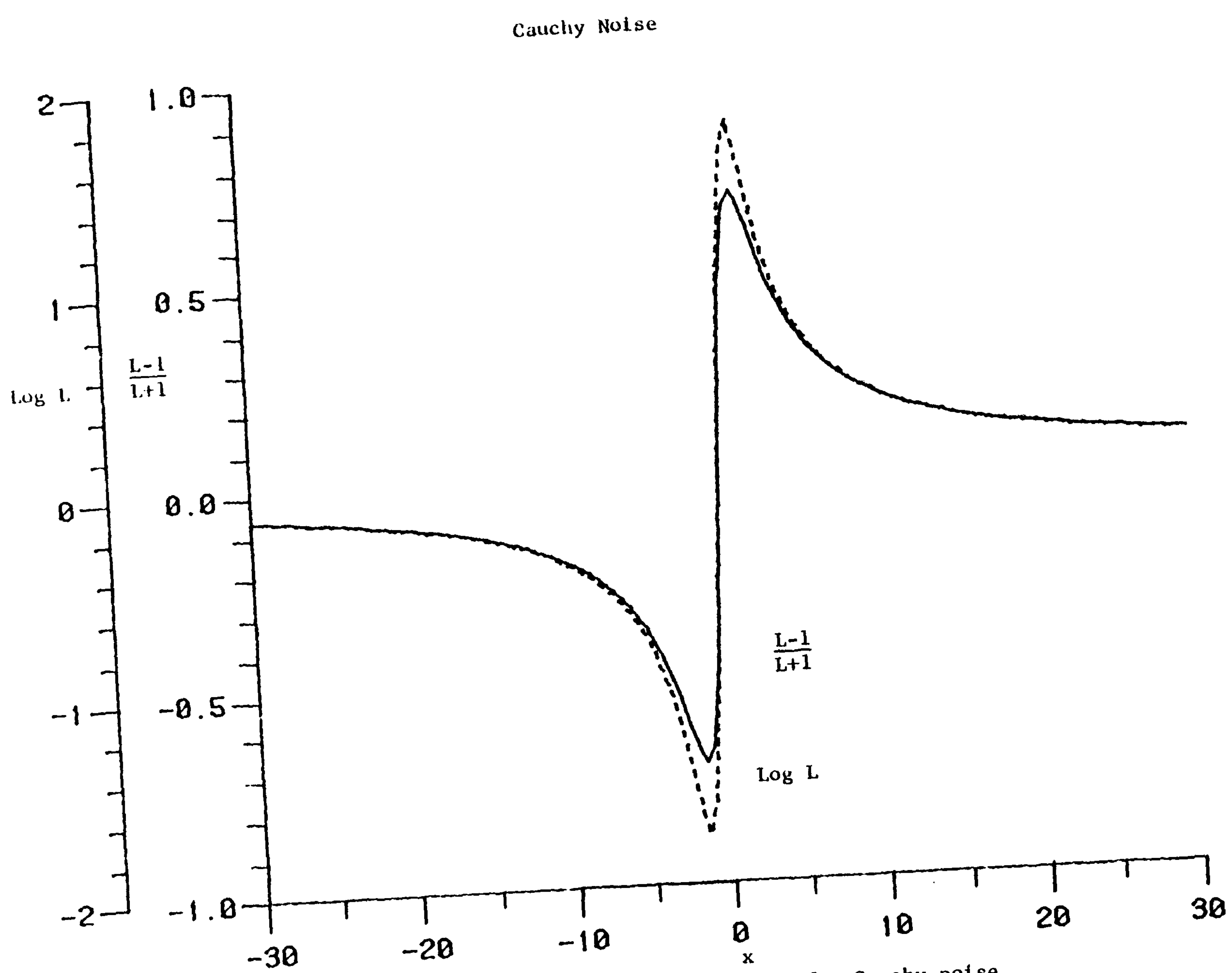


Figure 10. Nonlinearities for Cauchy noise.

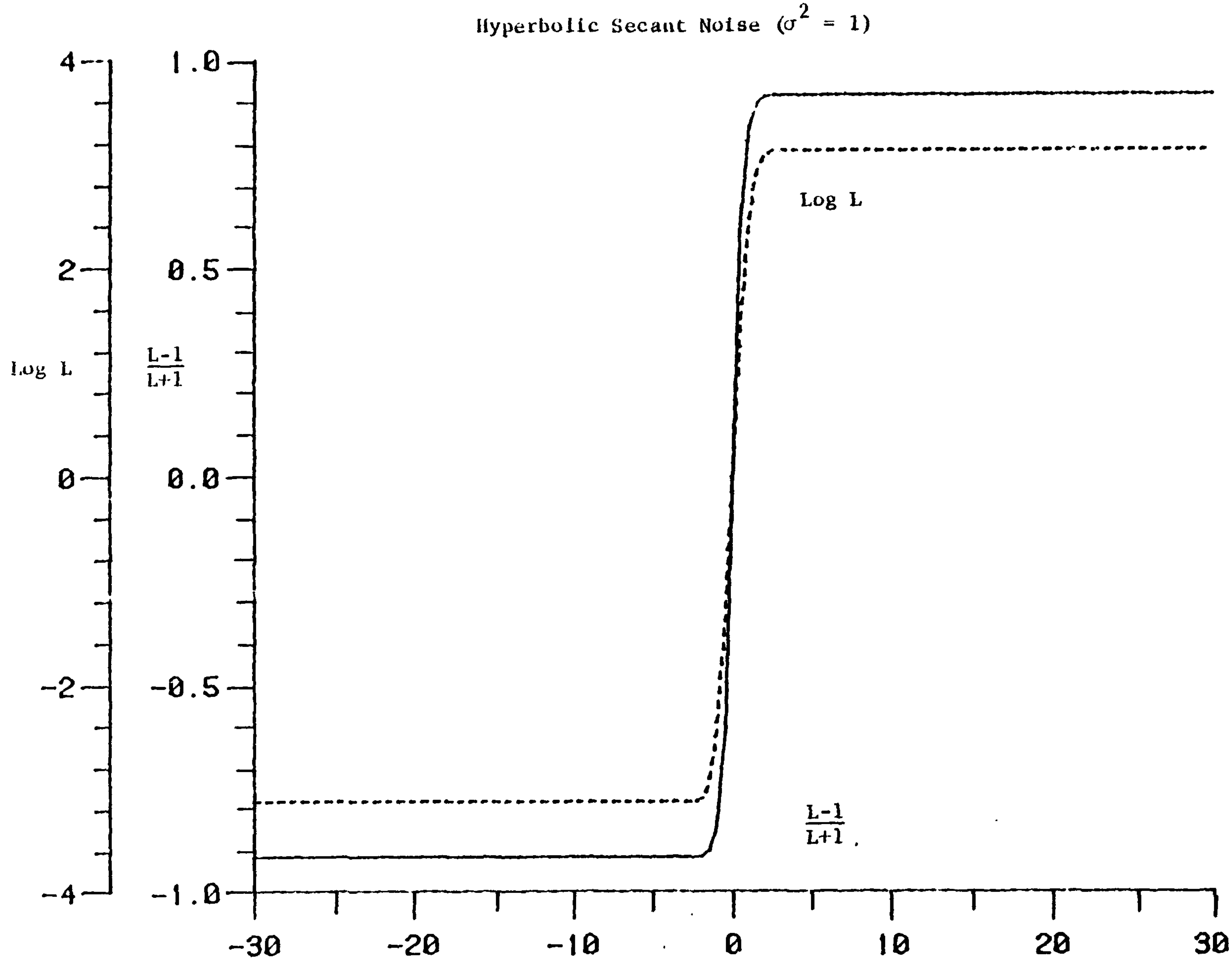


Figure 11. Nonlinearities for Sech noise.

TABLE I

Expected sample size for following noise densities under TANH($\frac{\text{LogL}}{2}$) and LogL. $\beta = 0.95$, signal = 0.4 and $\sigma^2 = 1$.

	Gaussian Noise		Cauchy Noise		Sech Noise	
α	LogL	TANH($\frac{\text{LogL}}{2}$)	LogL	TANH($\frac{\text{LogL}}{2}$)	LogL	TANH($\frac{\text{LogL}}{2}$)
1×10^{-1}	6.23	6.16	12.95	12.94	4.89	4.88
1×10^{-2}	13.05	12.89	27.12	27.11	10.25	10.24
1×10^{-3}	19.89	19.65	41.31	41.30	15.62	15.60
1×10^{-4}	26.72	26.41	55.51	55.50	20.98	20.96
1×10^{-5}	33.56	33.16	69.71	69.70	26.35	26.32
1×10^{-6}	40.39	39.92	83.91	83.90	31.72	31.69

TABLE II

Expected sample size for following noise densities under TANH($\frac{\text{LogL}}{2}$) and LogL. $\alpha = 10^{-5}$, $\beta = 0.95$ and $\sigma^2 = 1$.

	Gaussian Noise		Cauchy Noise		Sech Noise	
Signal	LogL	TANH($\frac{\text{LogL}}{2}$)	LogL	TANH($\frac{\text{LogL}}{2}$)	LogL	TANH($\frac{\text{LogL}}{2}$)
0.1	536.94	536.90	1076.56	1076.55	434.33	434.33
0.2	134.23	134.11	271.13	271.12	107.92	107.91
0.3	59.66	59.40	121.95	121.94	47.49	47.47
0.4	33.56	33.16	69.71	69.70	26.35	26.32
0.5	21.48	20.94	45.51	45.49	16.58	16.54
0.6	14.91	14.25	32.33	32.31	11.29	11.23
0.7	10.96	10.19	24.36	24.33	8.11	8.03
0.8	8.39	7.53	19.17	19.13	6.06	5.97
0.9	6.63	5.70	15.95	15.54	4.67	4.56
1.0	5.37	4.39	13.01	12.96	3.68	3.56
1.1	4.44	3.41	11.09	11.03	2.96	2.82
1.2	3.73	2.68	9.62	9.55	2.42	2.27
1.3	3.18	2.11	8.46	8.38	2.01	1.84
1.4	2.74	1.67	7.53	7.45	1.69	1.51
1.5	2.39	1.32	6.77	6.69	1.43	1.24

CHAPTER 4

NUMERICAL RESULTS

Consider our integral equation which is a Fredhelm integral equation of the second kind, i.e.

$$\hat{g}(x) - \int_0^\infty K(x,y)g(y)dy = g(x)$$

where $\hat{g}(x)$, $K(x,y)$ and $g(\cdot)$ are defined in Chapter 2 and $K(x,y)$ is the symmetric kernel. In general, if the kernel can be written in terms of some orthogonal functions on the interval $(-\infty,\infty)$, then the closed form solution can be easily obtained by the method of Hilbert-Schmidt (see Lovitt, 1950). Conversely, if such functions cannot be found, often a numerical procedure is used to solve this general equation. Consider the following numerical method.

4.1 General Procedure

A given operator equation

$$Pf = g \tag{4.1}$$

can be solved approximately for f by substituting

$$f \approx f_n \tag{4.2}$$

where f_n is a suitable approximating function depending upon n parameters $\alpha_1,\alpha_2,\ldots,\alpha_n$, substituting (4.2) in (4.1) to give

$$\delta_n = g - Pf_n \tag{4.3}$$

and then minimizing the residual function δ_n in some sense. Provided that this minimum δ_n^* is small enough, the function f_n^* determined by the optimal values α_j^* of the parameter in (4.3) usually provides a satisfactory approximation to the solution of (4.1).

This type of method can be applied to nonlinear operator equations, but most of our experience with it has been limited to the case of linear operator equations. In this case we solve

$$Lf = g \tag{4.4}$$

where L is a linear operator, by substituting

$$f \approx f_n = \sum_{j=1}^{n} \alpha_j \phi_j \tag{4.5}$$

in (4.4) and minimizing

$$\delta_n = g - Lf_n = g - \sum_{j=1}^{n} \alpha_j (L\phi_j)$$

$$= g - \sum_{j=1}^{n} \alpha_j \psi_j \quad . \tag{4.6}$$

Then we determine

$$\|\delta_n^*\| = \min_{\alpha_1, \alpha_2, \ldots, \alpha_n} \|g - \sum_{j=1}^{n} \alpha_j \psi_j\| \tag{4.7}$$

The parameter values α_j^* calculated in (4.7) give an approximate solution

$$f_n^* = \sum_{j=1}^{n} \alpha_j^* \phi_j$$

to (4.4). In practice we do not solve (4.7) exactly, but instead the simplex method of linear programming is used (Liu, 1968).

Example:

Consider the Fredholm integral equation of the second kind

$$f(x) - \frac{1}{4}\int_0^{\pi/2} xyf(y)dy = \sin x - \frac{1}{4}x, \quad 0 \le x \le \frac{\pi}{2} \tag{4.8}$$

for which the exact solution is $f(x) = \sin x$, suppose that the approximating function is chosen to be

$$f_3(x) = \alpha_1 x + \alpha_2 x^3 + \alpha_3 x^5 \tag{4.9}$$

then we have $n = 3$, $\phi_1 = x$, $\phi_2 = x^3$ and $\phi_3 = x^5$ substituting (4.9) in (4.8) yields

$$\delta_3(x) = g(x) - \sum_{j=1}^{3} \alpha_j \psi_j(x) \tag{4.10}$$

where

$$g(x) = \sin x - \frac{1}{4}x$$

$$\psi_1(x) = x - \frac{x}{4}\int_0^{\pi/2} y^2 dy = x(1 - \frac{1}{12}(\frac{\pi}{2})^3)$$

$$\psi_2(x) = x^3 - \frac{x}{4}\int_0^{\pi/2} y^4 dy = x(x^2 - \frac{1}{20}(\frac{\pi}{2})^5)$$

$$\psi_3(x) = x^5 - \frac{x}{4}\int_0^{\pi/2} y^6 dy = x(x^4 - \frac{1}{28}(\frac{\pi}{2})^7) \quad .$$

The next step is to minimize the residual $\delta_3(x)$ in (4.10). In practice we replace (4.10) by a discrete problem

$$\min_{\alpha_1,\alpha_2,\alpha_3} \max_{1 \le i \le m} \left|g(x_i) - \sum_{j=1}^{3} \alpha_j \psi_j(x_i)\right| \quad . \tag{4.11}$$

We often choose $m \approx 10\,n$, for this example the value of α_j^* obtain by solving (4.11) with $m = 33$ points, gives the approximate equation

$$f_3^*(x) = 0.99970x - 0.16567x^3 + 0.00751x^5$$

which is very close to sin x.

4.2 Example - Gaussian Noise

Consider the following zero-mean, unit-variance Bivariate Gaussian Noise density which in terms of Hermite polynomials

$$f_{N_1,N_j+1}(x,y) = f(x)f(y) \sum_{\nu=0}^{\infty} \frac{\rho_j^\nu He_\nu(x)He_\nu(y)}{\nu!} \tag{4.12}$$

where ρ_j is the coefficient of correlation and f denotes the standard Gaussian density. Here $He_\nu(\cdot)$ is the Hermite polynomial of order ν (note that, $H_\nu\left(\frac{-x}{\sqrt{2}}\right) = He_\nu(x)$). Substituting Eq. (4.12) in Eq. (2.50), after some arrangements we obtain,

$$\gamma \frac{L(x)-1}{L(x)+1} - mT(x) \int_{-\infty}^{\infty} \sum_{\nu=0}^{\infty} \sum_j \rho_j^\nu J_\nu(x)J_\nu(y)g(y)dy = g(x) \tag{4.13}$$

where

$$T(x) = \left[\sqrt{\frac{2}{\pi}} \exp\left\{- \frac{x^2+s^2}{2}\right\} \cdot \cos h(xs) \right]^{-1} \tag{4.14}$$

$$J_\nu(x) = [f(x-s)He_\nu(x-s) - f(x+s)He_\nu(x+s)] \tag{4.15}$$

and γ is an arbitrary nonzero constant. It can be easily shown that $J_\nu(\cdot)$ is not an orthogonal polynomial, therefore we apply the numerical method in Section (4.1) to obtain an approximate solution to Eq. (4.13). This is plotted in Fig. 12 for various values of m which is a parameter indicating the level of dependency. It is also interesting to know that the sample size required under these nonlinearities, and this is shown in Table III.

From the results, we may say that the sample size required in the m-dependent case is less than that needed in the independent case, and that the required sample size decreases as m increases. In other words, we might save some samples by increasing the sampling rate.

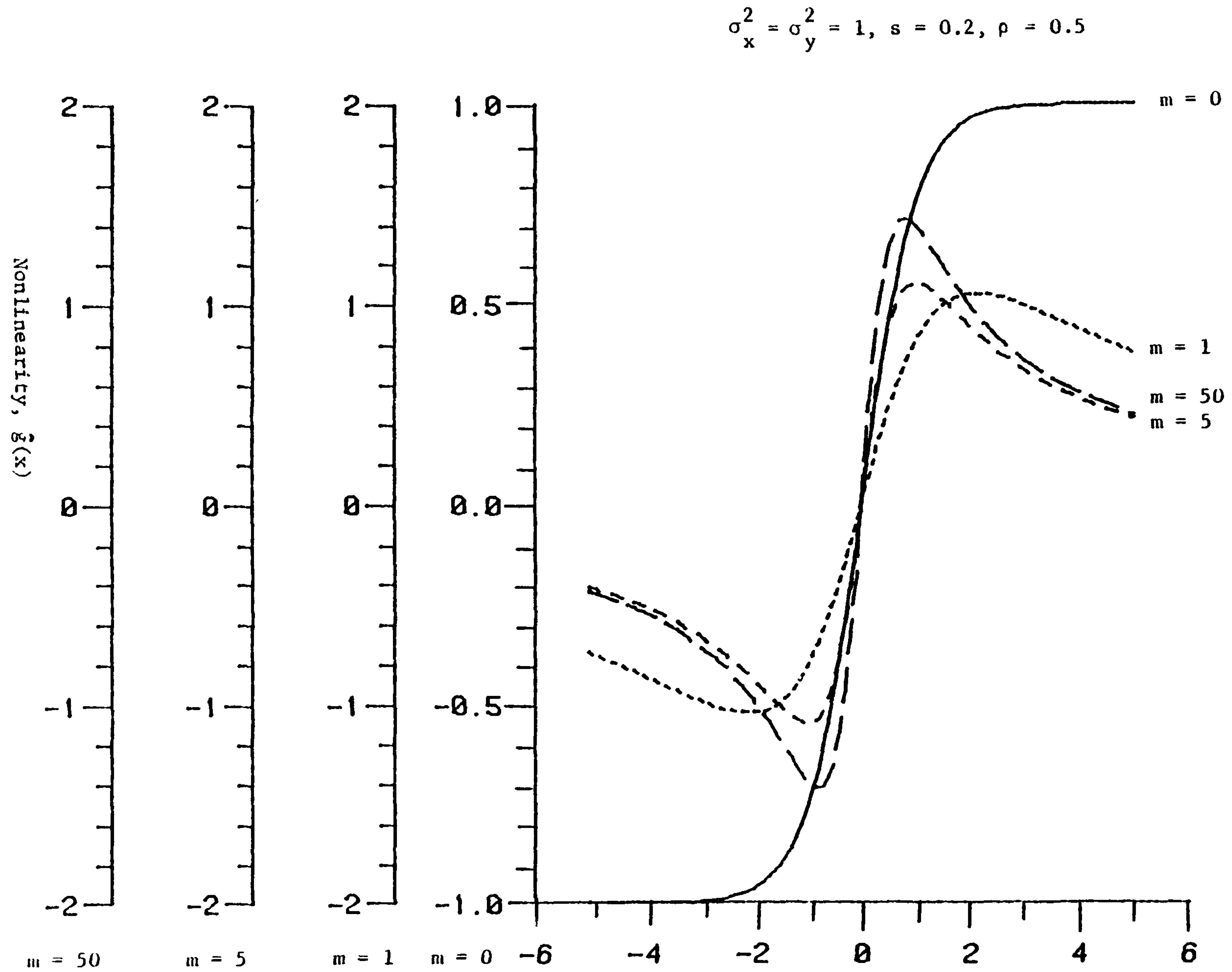


Figure 12. Nonlinearity function $\hat{g}(x)$ for different m.

Table III

Expected sample size for Gaussian noise density

$\alpha = 10^{-5}$, $\beta = 0.95$, $\sigma_1^2 = \sigma_2^2 = 1$, $\rho = 0.5$

Signal	m = 1	m = 5	m = 50
0.1	513.78	483.69	443.72
0.2	128.33	120.82	110.83
0.3	56.84	53.51	49.10
0.4	31.73	29.87	27.40
0.5	20.01	18.86	17.31

CHAPTER 5

CONCLUSIONS

This thesis has considered several aspects of sequential detection systems. It has been shown that the sample size needed for the case of m-dependent process is less than the case of independent process required. A comparison between LogL and TANH(LogL/2) is also examined for the asymptotic case, and it has shown that TANH(LogL/2) indeed is the optimum nonlinearity under our criterion. Finally, for the Gaussian noise density, we compute the sample size and plotted the nonlinearities under different signals and level of dependencies (m). For further investigation one might study the ARE between m-dependent sequential detector and m-dependent fixed sample detector.

APPENDIX

To show $J_g(\epsilon) = J_g(0) + \epsilon J_g'(0) + \lambda\epsilon^2\sigma_1^2(\delta g)$

From Eq. (2.37) we have

$$J_g(\epsilon) = H(g + \epsilon\delta g)$$

$$= \int (g + \epsilon\delta g)f_1 + \lambda\sigma_1^2(\epsilon\delta g + g)$$

$$= \int gf_1 + \lambda\sigma_1^2(g) - \lambda\sigma_1^2(g) + \int \epsilon\delta g f_1 + \lambda\sigma_1^2(\epsilon\delta g + g)$$

$$= J_g(0) - \lambda\sigma_1^2(g) + \int \epsilon\delta g f_1 + \lambda\sigma_1^2(\epsilon\delta g + g)$$

$$= J_g(0) + \epsilon J_g'(0) - \epsilon\lambda[2\int g\delta g f_1 + 2\sum_{j=1}^{m}\iint 2$$

$$\cdot\ g\delta g f_{N_1,N_{j+1}} - 2(2m+1)(\int gf_1)(\int \delta g f_1)] - \lambda\sigma_1^2(g) + \lambda\sigma_1^2(\epsilon\delta g + g)$$

$$= J_g(0) + \epsilon J_g'(0) + \epsilon^2\lambda\{\int (\delta g)^2 f_1 + 2\sum_{j=1}^{m}\iint (\delta g)^2 f_{N_1,N_{j+1}}$$

$$- (2m+1)(\int gf_1)^2\}$$

$$= J_g(0) + \epsilon J_g'(0) + \epsilon^2\lambda\sigma_1^2(\delta g)$$

where $f_{N_1,N_{j+1}}$ is the joint density of N_1 and N_{j+1} and f_1 is the density function under H_1.

REFERENCES

T. W. Anderson, "A modification of the sequential probability ratio test to reduce the sample size," Ann. Math. Stat., vol. 31, pp. 165-197, 1960.

P. M. Anselone, Collective Compact Operator Approximation Theory, Prentice-Hall, 1971.

P. Beckmann, Orthogonal Polynomials for Engineers and Physicists, Colem, 1973.

W. H. Beyer, CRC Standard Math Tables, 25th edition, CRC Press, 1978.

P. Billingsley, Convergence of Probability Measures, Wiley, 1968. Chapter 4.

J. J. Bussgang, "Sequential methods in radar detection," Proc. IEEE, vol. 58, pp. 731-743, May 1970.

T. S. Ferguson, Mathematical Statistics, Academic, 1967.

A. A. Galvin, "A sequential detection system for the processing of radar returns," IRE, pp. 1417-1423, 1961.

B. K. Ghosh, Sequential Tests for Statistical Hypothesis, Addison-Wesley, 1970.

Z. Govindarajulu, Sequential Statistical Procedures, Academic Press, 1975.

R. V. Hogg and A. T. Craig, Introduction to Mathematical Statistics, Third edition, MacMillan, 1970.

H. H. Kagiwada and R. Kalaba, Integral Equations via Imbedding Methods, Addison-Wesley, 1974.

J. Kiefer and L. Weiss, "Some properties of generalized sequential probability ratio tests," Ann. Math. Stat., vol. 34, pp. 57-74, 1957.

T. L. Lai, "Optimal stopping and sequential tests which minimize the maximum expected sample size," Ann. Stat., vol. 1, pp. 659-673, July 1973.

C. Liu, Introduction to Combinatorial Mathematics, McGraw-Hill, 1968.

W. V. Lovitt, Linear Integral Equations, Dover, 1950.

A. Papoulis, Probability Random Variables and Stochastics Processes, McGraw-Hill, 1965.

W. Pogorzelski, Integral Equations and Their Applications, vol. 1, Pergamon, 1966.

H. V. Poor and J. B. Thomas, "Asymptotically optimum zero-memory detectors for m-dependent noise processes," Proc. 1977 Johns Hopkins Conf. on Inform. Sciences and Systems, pp. 134-139, 1977.

H. V. Poor and J. B. Thomas, "Memoryless discrete-time detection of a constant signal in m-dependent noise," IEEE Trans. on IT, vol. IT-25, no. 1, 1979.

M. R. Reynolds, Jr., "A sequential signed-rank test for symmetry," Ann. of Stat., vol. 3, pp. 382-400, 1975.

P. K. Sen and M. Ghosh, "Sequential rank tests for location," Ann. of Stat., vol. 1.2, pp. 540-552, 1974.

S. Tantaratana, On the Relative Efficiencies of Some Parametric and Nonparametric Sequential Detectors, Ph.D. Dissertation, Princeton University, 1977.

F. G. Tricomi, Integral Equations, Interscience, 1957.

H. Van Tree, Detection, Estimation, and Modulation Theory, Wiley, 1968.

A. Wald, Sequential Analysis, Wiley, 1947.

G. L. Wise, The Bivariate Structure and Representation of Random Processes, Ph.D. Dissertation, Princeton University, 1974.

P. P. Zabreyko et al., Integral Equations, Noordhovv, 1975.

3-8

DT